

col - H02799 - 5 - p002818

AMERICAN PHILOSOPHICAL QUARTERLY

5

Edited by
NICHOLAS RESCHER

With the advice and assistance of the Board of Editorial Consultants:

Alan R. Anderson	Adolf Grünbaum	Richard Rorty
Kurt Baier	Carl G. Hempel	George A. Schrader
Monroe Beardsley	John Hospers	Michael Scriven
Nuel D. Belnap, Jr.	Raymond Klibansky	Wilfrid Sellars
Roderick M. Chisholm	Hugues Leblanc	Alexander Sesonske
L. Jonathan Cohen	Ernan McMullin	J. J. C. Smart
James Collins	Benson Mates	Manley H. Thompson, Jr.
James M. Edie	John A. Passmore	James F. Thomson
Peter Thomas Geach	Günther Patzig	John W. Yolton
	Richard H. Popkin	

VOLUME 5/NUMBER 3

JULY 1968

CONTENTS

I. ALVIN I. GOLDMAN: <i>Actions, Predictions, and Books of Life</i> . . .	135	V. JOHN L. POLLOCK: <i>What Is an Epistemological Problem?</i> . . .	183
II. C. A. HOOKER: <i>Craigian Transcriptionism</i> . . .	152	VI. JON MOLINE: <i>On Points of View</i> . . .	191
III. GILBERT HARMAN: <i>Knowledge, Inference, and Explanation</i> . . .	164	VII. JON WHEATLEY: <i>Evaluation</i> . . .	199
IV. SID B. THOMAS, Jr.: <i>The Status of The Generalization Principle</i> . . .	174	VIII. JOSEPH MARGOLIS: <i>On Names: Sense and Reference</i> . . .	206
		IX. HERBERT HEIDELBERGER: <i>The Indispensability of Truth</i> . . .	212

PUBLISHED BY BASIL BLACKWELL WITH THE COOPERATION OF THE UNIVERSITY OF PITTSBURGH

AMERICAN PHILOSOPHICAL QUARTERLY

POLICY

The *American Philosophical Quarterly* welcomes articles by philosophers of any country on any aspect of philosophy, substantive or historical. However, only self-sufficient articles will be published, and not news items, book reviews, critical notices, or "discussion notes."

MANUSCRIPTS

Contributions may be as short as 2,000 words or as long as 25,000. All manuscripts should be typewritten with wide margins, and at least double spacing between the lines. Footnotes should be used sparingly and should be numbered consecutively. They should also be typed with wide margins and double spacing. The original copy, not a carbon, should be submitted; authors should always retain at least one copy of their articles.

COMMUNICATIONS

Articles for publication, and all other editorial communications and enquiries, should be addressed to: The Editor, *American Philosophical Quarterly*, Department of Philosophy, University of Pittsburgh, Pittsburgh, Pennsylvania 15213.

OFFPRINTS

Authors will receive gratis two copies of the issue containing their contribution. Offprints can be purchased at low cost through arrangements made when checking proof.

SUBSCRIPTIONS

The price *per annum* is six dollars for individual subscribers and ten dollars for institutions. Checks and money orders should be made payable to the *American Philosophical Quarterly*. All back issues are available, and are sold at the rate of two dollars to individuals, and three dollars to institutions. All correspondence regarding subscriptions and back orders should be addressed directly to the publisher (Basil Blackwell, Broad Street, Oxford, England).

MONOGRAPH SERIES

The *American Philosophical Quarterly* also publishes a supplementary Monograph Series. Apart from the publication of original scholarly monographs, this includes occasional collections of original articles on a common theme. These monographs can be purchased separately but are available to individual subscribers at a substantially reduced price. The back cover of the journal may be consulted for details.



I. ACTIONS, PREDICTIONS, AND BOOKS OF LIFE

ALVIN I. GOLDMAN

ARE actions determined? Since it is difficult to tell "directly" whether or not actions are governed by universal laws, some philosophers resort to the following "indirect" argument:

If actions are determined, then it is possible in principle to predict them (with certainty).

It is not possible in principle for actions to be predicted (with certainty).

Therefore, actions are not determined.

A defender of this argument I shall call an "anti-predictionist"; his position will be called "anti-predictionism." In this paper I shall try to rebut anti-predictionism.

Both premisses of the anti-predictionist argument will come under attack here. The first premiss, affirming that determinism entails predictability, is often accepted without adequate scrutiny. Some writers not only assume that determinism entails predictability but even *define* determinism as the thesis that every event is predictable in principle.¹ I believe, however, that it is essential to distinguish between determinism and predictability. We must first notice that there are various kinds or senses of "possibility" which may be involved in the "possibility of prediction." Moreover, it can be shown that in many of these senses, determinism does *not* entail the possibility of prediction. Many anti-predictionists have failed to notice this, however. Therefore, upon discovering some unpredictability in the arena of human action, they have wrongly concluded that actions must be undetermined. This error will be avoided only if we carefully distinguish between determinism and predictability. Hence, an important aim of this paper will be to differentiate various senses of "possibility of prediction" and to ascertain how they are related to determinism.

Let us assume now that we can find some suitable sense of "possibility of prediction" which is closely related to, if not entailed by, determinism. The second premiss of the anti-predictionist argument

asserts that, in such a sense, it is impossible for actions to be predicted. Various arguments have been offered in support of this premiss. One that I shall consider concerns the possibility of writing a complete description of an agent's life—including his voluntary actions—even before he is born. According to anti-predictionism, if actions were determined, it would be possible to write such books. Indeed, it would be possible for such a "book of life" to be written even if the agent were to read its prediction of a given action before he is to perform that action. It seems clear to the anti-predictionist, however, that such books of life are impossible. Predictions of my actions cannot be made with certainty; for when I read these predictions, I can easily choose to falsify them. So argues the anti-predictionist. But it is far from clear that he is right. I think, on the contrary, that it may well be possible (in a suitable sense) for books of life to be written. And thus it seems to me that the anti-predictionist is unable to establish the truth of his second premiss.

In general, anti-predictionists support their second premiss by contrasting the predictability of human behavior with that of physical events. It is alleged that special difficulties of a purely conceptual sort arise for the prediction of action and that these difficulties are unparalleled in the realm of merely physical phenomena. I shall claim, however, that there are no essential differences between actions and physical events with respect to the problem of prediction. More precisely, I shall claim that *conceptual* reflection on the nature of human behavior (as opposed to investigation by the special sciences) does not reveal any peculiar immunity to prediction.

It must be emphasized that I offer no proof of the thesis that actions *are* determined; I merely wish to show that the anti-predictionist's arguments fail to prove that they are *not* determined. It is conceivable, of course, that actions are not determined. And if actions are not determined, then I would admit that they are not perfectly

¹ Karl Popper, for example, defines "determined" as "predictable in accordance with the methods of science," in "Indeterminism in Quantum Mechanics and in Classical Physics," *The British Journal for the Philosophy of Science*, vol. I (1950-51), see p. 120.

predictable (in any sense at all). What I contend, however, is that the arguments of philosophers, based on familiar, common-sense features of human action and human choice, do not prove that actions are undetermined or unpredictable. The basic features of human action are quite compatible with the contention that actions are determined and susceptible of prediction. In other words, my aim here is not to establish the *truth*, but merely the *tenability*, of the thesis that actions are determined.

II

Let us begin with some definitions. I shall define determinism as the view that every event and state of affairs is determined in every detail. An event is determined (in a given detail) if and only if it is deducible from some set of antecedent conditions and laws of nature. A law of nature is, roughly, any true non-analytic universal statement of unlimited scope which supports counterfactual conditionals.² Both "low-level" empirical connections, like all metals expand when heated, and "theoretical" connections, like $F = ma$, are included. Antecedent conditions can be either events, like moving at 10 m.p.h., or states of affairs, like having a specific gravity of 1.7. (Throughout I shall be concerned both with events and with states of affairs, but for brevity I shall often omit reference to states of affairs.) Negations of events, like a ball's *not* moving at 10 m.p.h., are also included. Antecedent conditions may be directly observable phenomena, but they need not be. Theoretical, hypothetical, and dispositional states—like being brittle or being intelligent—can serve as antecedent conditions.

Notice that my definition of determinism is in terms of a formal relationship, i.e., the relationship of deducibility holding between events and sets of laws and antecedent conditions. In particular, this definition makes no explicit reference to the ability of anyone to predict these events, and thereby leaves open the question of the connection between determinism and predictability.

If determinism is true, human actions are determined. But determinism alone does not tell us what laws or kinds of laws take human actions as their dependent variables. I shall assume, however,

that these laws would include ones with psychological states like desires, beliefs, intentions, etc., as their independent variables. This presupposes—correctly, I think—that statements connecting actions with, for example, wants and beliefs, are not purely analytic.³ Rather, their logical status would correspond to quasi-analytic, quasi-empirical generalizations like many theoretical statements of science. If determinism is true, wants, beliefs, intentions, etc., are themselves determined by prior events of various sorts. The determinants of these mental states are quite diverse, however, so I shall make no attempt to delineate them.

In ordinary language, not all determining factors of an event are called its "causes." A body's having a certain mass may be a determining antecedent condition of that body's moving at a certain velocity after being struck by another object, but its having that mass would not be called a "cause" of its velocity. Similarly, although a person's having a certain intention or desire would not ordinarily be termed a "cause" of his action, it may be an antecedent condition of the relevant sort. Since determinism is often connected with what philosophers call "causal necessity," I shall use the technical term "*causally necessitate*" to apply to antecedent conditions which, together with laws of nature, determine a given event. Thus, I shall say that desires and beliefs (together with other conditions) "*causally necessitate*" a given action, even though ordinary language would not condone such an expression.

In our discussion of predictability we need a sense of "prediction" distinct from mere lucky guesses or pre-cognition. We must be concerned with predictions made on the basis of laws and antecedent conditions. I shall call a prediction a "*scientific prediction*" if and only if it is made by deducing the predicted event from laws and antecedent conditions. A scientific predictor may learn of the laws and antecedent conditions in any number of ways. (On my definition, most predictions made by actual scientists are not "scientific" predictions, for real scientists seldom, if ever, *deduce* what will occur from laws and prior conditions. Nevertheless, scientific prediction as defined here may be regarded as an ideal of prediction to which scientists can aspire.)

² There are, of course, numerous problems associated with the concept of a law of nature. But a detailed discussion of these problems would go beyond the scope of this paper.

³ For a defense of this view, see William P. Alston, "Wants, Actions, and Causal Explanations" in H. N. Castañeda, (ed), *Minds, Intentionality, and Perception* (Detroit, 1967) and R. Brandt and J. Kim, "Wants as Explanations of Actions," *The Journal of Philosophy*, vol. 60 (1963), pp. 425-435.

As indicated above, it is important to identify different senses of the phrase "possibility of prediction." I shall now distinguish four relevant species of possibility, though further distinctions will be made later within some of these categories. The four species are: (1) *logical possibility*, (2) *logical compossibility*, (3) *physical possibility*, and (4) *causal compossibility*.

An event is *logically possible* if and only if it is not self-contradictory, and logically impossible if and only if it is self-contradictory. Drawing a square circle is a logically impossible event, while jumping 90 feet is a logically possible event. *Logical compossibility* is defined for two or more events. A set of two or more events is logically compossible if and only if the conjunction of the members of the set is logically consistent. A set is logically impossible (i.e., not logically compossible) if and only if each of the events is logically possible but their conjunction is logically inconsistent. Thus, the two events, (a) x 's being a pumpkin from 11 o'clock to 12 o'clock, and (b) x 's turning into a pumpkin at 12 o'clock, are logically impossible.

An event is *physically possible* if and only if it is not inconsistent with any law or laws of nature; an event is physically impossible if and only if there are laws of nature with which it is inconsistent. Traveling faster than the speed of light, for example, is physically impossible. I shall speak not only of events being physically impossible *in general*, but also as being physically impossible *for* certain kinds of entities. Thus, the act of lifting a ten-ton weight is not, in general, physically impossible; but it is physically impossible for (normal) human beings to lift ten-ton weights. Given the physical constitution of human beings, laws of nature make it impossible for them to lift such weights.

Causal compossibility differs from physical possibility in attending to groups of events rather than events taken singly. Roughly, a set of events is causally compossible just in case laws of nature allow each of them to occur singly and allow them to occur as a group. More precisely, consider a set of events $\{e_1, \dots, e_n\}$ each of which is logically possible and physically possible, and which are jointly logically compossible. Then $\{e_1, \dots, e_n\}$ is a causally compossible set if and only if there is

no set of laws of nature such that the conjunction of these laws with e_1, \dots, e_n is logically inconsistent.⁴ I shall say that the set as a whole is causally compossible or that each member is causally compossible "with" or "relative to" the other members.

A set of events $\{e_1, \dots, e_n\}$ is causally impossible (i.e., not causally compossible) if and only if there are some laws of nature L_1, \dots, L_k such that the conjunction of L_1, \dots, L_k with e_1, \dots, e_n is logically inconsistent. Assuming, as we do, that e_1, \dots, e_n satisfy the other three species of possibility, the set $\{e_1, \dots, e_n\}$ will be causally impossible if and only if the negation of (at least) one member of the set is entailed by the conjunction of the other members of the set conjoined with L_1, \dots, L_k . Thus, if the negation of a given member of the set is causally necessitated by the other members of the set, then the set is causally impossible.

III

The most interesting questions concerning the prediction of action are best handled in terms of the notion of causal compossibility. The reflexivity of predictions—the fact that a prediction often has an effect which bears on its own truth—can be understood properly with the use of this notion. But the question of the causal compossibility of predictions of action cannot arise unless the other three species of possibility are satisfied. Our definition of causal compossibility makes a set causally compossible only if its members are logically possible, physically possible, and (jointly) logically compossible. For example, if it is physically impossible to make scientific predictions of actions, the question of causal compossibility does not even arise. Therefore, before turning to the questions of reflexivity, including the question of whether "books of life" can be written, we must focus on certain problems connected with the logical compossibility and the physical possibility of predicting actions.

The logical possibility and compossibility of predictions can be discussed together, since the distinction between them is somewhat blurred. This is because a correct prediction is not really a single event, but a pair of events—a prediction and an

⁴ The term "event" is here used to designate event *kinds*, not necessarily ones that have been actualized. The term "law," on the other hand, will be used only to designate actual laws, i.e., laws that obtain in the real world, and not merely possible laws.

For the most part, I shall consider events with built-in time references. Sam's jumping rope at 10:35 will be treated as a distinct event from Sam's jumping rope at 10:45. This is very natural in the present context, since a given set of events may be causally impossible with Sam's jumping rope at 10:35 but causally compossible with Sam's jumping rope at 10:45.

event predicted. Two different examples of logical impossibility have been uncovered in connection with the prediction of behavior. I shall discuss these examples briefly and argue that, contrary to what their authors suppose, they do not prove that actions are undetermined and they do not prove that actions have a peculiar immunity to prediction unparalleled by physical phenomena.

The first logical impossibility, as discussed by Maurice Cranston,⁵ can be summarized as follows. Suppose that Sam invents the corkscrew at time t . In the intended sense of "invent," this means (a) that Sam thinks of the corkscrew at t , and (b) that no one ever thought of the corkscrew before t . Cranston argues that no one could have predicted Sam's inventing the corkscrew. In order for him to make this prediction, he would himself have to think of the corkscrew. And had he thought of the corkscrew, it would be false to say that Sam "invented" the corkscrew. Yet, *ex-hypothesi*, Sam *did* invent the corkscrew. Using the terminology of "logical impossibility," we can formulate Cranston's problem by saying that the three events, (a) Sam thinks of the corkscrew at t , (b) no one ever thought of the corkscrew before t , and (c) someone predicted Sam's inventing the corkscrew, are logically impossible.

The second example poses a problem for predicting not actions, but decisions. However, since the concept of a voluntary action is so closely tied to that of a decision, an unpredictability connected with decisions is very important for us to discuss. Carl Ginet claims that it is impossible ("conceptually" impossible) for anyone to predict his own decisions.⁶ The argument begins by defining "deciding to do A " as *passing into* a state of knowledge (of a certain kind) that one will do A , or try to do A .⁷ Suppose now that Sam, at t , decides to do A . Had Sam predicted that he would make this decision—and had this prediction involved *knowledge*—he could not have decided later to do A . For if, before t , he had known that he would decide to do A , he would have known then that he would do A , or try to do A . But if, before t , he had known that he would do A (or try to do A), then he

could not, at t , have *passed into* a state of knowing that he would do A . Thus, according to Ginet, Sam could not have predicted that he would make this decision.

Of course, Sam might make his prediction and then forget it. If so, he can still decide, at t , to do A . However, if Sam not only knows, before t , that he will decide to do A , but also *continues* to know this up until t , then Sam cannot, at t , decide to do A . In other words, the following three events are logically impossible: (a') Sam decides, at t , to do A , (b') Sam predicts (i.e., knows) that he will decide to do A , and (c') Sam continues to know this until t .

What do these two logical impossibilities prove? Do they prove that decisions and inventions are undetermined? Do they prove that voluntary actions, including the decisions which lead to them, have a special immunity to prediction? The answer is "No," I believe, to both questions.

Our examples of logical impossibilities do not establish any special status for human behavior, for precisely analogous impossibilities can be produced for physical phenomena. Let the expression "a tornado strikes x by surprise" mean: (1) a tornado strikes x at a certain time, and (2) before that time nobody ever thought of a tornado striking x . Now suppose that, as a matter of fact, a tornado strikes Peking by surprise. Then it is logically impossible for this event to have been predicted. That is, the set consisting in the tornado striking Peking by surprise and a prediction of the tornado striking Peking by surprise is a logically impossible set. In general it is logically impossible for tornadoes striking places by surprise to be predicted. For if anyone were to predict these events, they could no longer be described as "tornadoes striking places by surprise." Nevertheless, there certainly are (or could be) events correctly describable as "a tornado striking x by surprise."

I wish next to argue that the invention and decision impossibilities do not show that these human phenomena are undetermined. Notice first that the tornado case, though it has the same

⁵ *Freedom: A New Analysis* (London, 1954), p. 169.

⁶ "Can the Will Be Caused?," *The Philosophical Review*, vol. 71 (1962), pp. 49-55.

⁷ One might challenge Ginet's argument by criticizing this definition of "deciding." This criticism has implicitly been made, along with other criticisms of Ginet's position, by various writers. For example, see John Canfield, "Knowing about Future Decisions," *Analysis*, vol. 22 (1962), and J. W. Roxbee Cox, "Can I Know Beforehand What I Am Going to Decide?," *The Philosophical Review*, vol. 72 (1963). Here I shall waive these criticisms, however, and accept Ginet's claim that it is impossible to predict one's own decisions. I shall then ask whether this proves that decisions are undetermined and whether they are intrinsically different from physical phenomena.

logical structure, does not bear on the question of determinism. Although it is logically impossible for anyone to predict the tornado striking Peking by surprise, I am in no way inclined to suppose that this event is not determined. Similarly, our logical impossibilities fail to show that inventions and decisions are undetermined. How could such logical impossibilities demonstrate that these events are not governed by laws of nature? The notion of a law is in no way involved in the concept of logical impossibility. And hence the presence of logical impossibilities sheds no light on the question of whether there are laws and antecedent conditions which entail inventions or decisions.

The critical error here is the assumption that if an event is determined (under a given description), it must be possible to predict it (under that description).⁸ The falsity of this proposition should be adequately clear from the invention case. Suppose that Sam's thinking of the corkscrew at t is deducible from laws and antecedent conditions. And suppose that the fact that no one ever thinks of the corkscrew before t is also deducible from laws and antecedent conditions. Then, the event consisting in Sam's *inventing* the corkscrew at t would be determined; but it still would be logically impossible for it to have been predicted under that description. The lesson to be learned here is not that inventions are undetermined actions, but that the alleged entailment between determinism and predictability is not an entailment at all. At any rate, the fact that an event is determined under a given description does not entail that it is *logically compossible* for it to be predicted under that description.⁹

The case of decisions can be handled similarly. It seems to me quite possible that a person's passing into a state of knowing, or intending, to do A be deducible from laws and antecedent conditions. But although this event would be determined (under the given description) it would not be logically compossible for Sam to have predicted it (under that description) and continued to know it until t .

⁸ That this is an error has also been claimed by Arnold S. Kaufman, in "Practical Decision," *Mind*, vol. 75 (1966), see p. 29.

⁹ It is also an error—committed at least as frequently—to think that determinism entails the possibility of retrodicting or explaining every event under any description. Suppose that Sam thinks of the corkscrew at t and that no one ever thinks of the corkscrew after t . Suppose, moreover, that both of these events are deducible from laws and antecedent conditions. Now let us introduce the expression "postventing x " to mean "thinking of x for the last time" (just as "inventing x " means "thinking of x for the first time"). Clearly, we may say of Sam that he "postvented" the corkscrew and that this action of his is determined. However, it is logically impossible for anyone to *retrodict* Sam's postventing the corkscrew. To do so, the retrodictor would himself have to think of the corkscrew, and, *ex hypothesi*, Sam thought of the corkscrew for the last time at t .

¹⁰ *Op. cit.*

IV

I turn now to physical possibility. Is it physically possible to make scientific predictions of human actions? Here the emphasis should be placed on the qualifier "scientific." Although it may well be physically possible to make "lucky guess" predictions, or perhaps even predictions based on "intuition," it is not obvious that predictions can be made by deducing an action from laws and antecedent conditions. And this is the only kind of prediction which bears on the issue of determinism.

Anti-predictionists might claim that it is physically impossible for human beings to make scientific predictions of actions, because human beings cannot learn enough antecedent conditions to deduce what will be done. But it is inessential to the predictionist's position to restrict the range of predictors to human beings. In order to avoid theological or supernatural issues, we may require that any predictor be a finite entity operating within the causal order of the universe. But apart from this, no arbitrary limits should be placed on admissible predictors.

Karl Popper¹⁰ has tried to show that there are certain limitations of the predictions which can be made by "classical mechanical calculating" machines. But to restrict the range of predictors to calculating machines is an important restriction; even if Popper is right about the prediction-limitations of machines of the sort he discusses, there may be other beings that can make predictions his machines cannot. Another limitation on Popper's discussion is that much of it is aimed at establishing the physical impossibility of a *single* being, like Laplace's demon, making scientific predictions of *all* events or of a very large number of events. But the fact that all events cannot be predicted by a single being is compatible with the proposition that every event can be predicted by some being or other.

Anti-predictionists might proffer the following arguments for saying that it is physically impossible for *any* finite being, not just human beings, to make

scientific predictions of human behavior. Scientific predictions, they might claim, require knowledge of infinitely many facts, but it is physically impossible for a finite being to know infinitely many facts. The infinity requirement seems necessary because in order to *deduce* that even a certain finite system will yield a given result, one must know that no interfering factors will intrude from outside the system. And knowing this may involve knowing *all* states of the world at least at one time.

This argument is of questionable force. It is far from clear that the deduction of actions from antecedent conditions and laws requires knowledge of infinitely many facts. Nor is it clear that no finite being could know infinitely many facts. Even if the argument is correct, however, it would seem to prove *too much*. For if the knowledge of infinitely many facts is required in order to make scientific predictions of actions, the same would be true for scientific predictions of physical events. Thus, the above argument would fail to establish any special immunity of human action to prediction. Finally, even if it is physically impossible for any finite being to make scientific predictions of actions, this would not prove that actions are undetermined. Here too, as above, we have a sense of "possibility" in which determinism does *not* entail the possibility of prediction. The proposition that an event is (formally) deducible from laws and antecedent conditions does not entail that it is physically possible for any being to come to know these laws and antecedent conditions and to deduce the event from them. Hence, even if the anti-predictionist could establish that it is physically impossible to predict actions scientifically, he would not thereby establish that actions are undetermined.

We have not conclusively shown either that it is physically possible for some beings to predict actions scientifically or that it is not. But unless we assume that this is physically possible, we cannot turn to the other interesting issues that surround the problem of the prediction of human behavior. Unless we assume this, the question of the causal compossibility of predicting actions cannot even arise. In order to explore these important issues, therefore, I shall henceforth assume that scientific predictions of actions (like scientific predictions generally) are physically possible.

V

Perhaps the anti-predictionist would think it obvious that it is causally impossible to predict

actions scientifically. He might argue as follows: "Let us grant, as is likely, that there have never been any genuine scientific predictions of voluntary human actions. If, as my opponent claims, determinism is true, then it is causally impossible for any predictions to have been made of these actions. For every actual action A , there is an actual event \bar{P}_A , the *absence* of a prediction of A . Since each of these events \bar{P}_A is actual, and since determinism is true, each of these events \bar{P}_A must be causally necessitated by some set of actual events prior to it. But if each of these events \bar{P}_A is causally necessitated by actual prior events, then each event \bar{P}_A —the prediction of A —is causally impossible relative to some actual events. In other words, for each actual action A , it is causally impossible for A to have been predicted."

This argument, like a previous one, proves too much. The anti-predictionist is right in saying that non-actual predictions of actions are causally impossible with the actual prior events in the world. But this is true simply because, assuming determinism, every non-actual event whatever is causally impossible with some set of actual prior events. Thus, using the notion of causal-compossibility-relative-to-all-actual-events, we can establish the impossibility of predicting physical phenomena as well as human behavior. We can point to an action that was never predicted and say that, in this sense, it "could not" have been predicted, since its non-prediction was causally necessitated by other actual events. But by the same token, we can point to a physical event which was never predicted and say that it "could not" have been predicted, since its non-prediction was also causally necessitated by other actual events. Using this notion of "possibility of prediction," the anti-predictionist again fails to establish any special immunity of action to prediction.

Apart from this point, however, the notion of "causal-compossibility-relative-to-all-actual-events" does not seem to be a pertinent kind of possibility for our discussion. We have seen that determinism does not entail the possibility of predicting actions in *every* sense of "possible." And here, I believe, we have still another sense of "possible" in which determinism does not entail that it is possible for every action to be predicted. Determinism does not say that, relative to all actual prior events, it is causally compossible for a prediction of an action to be made *even if* those actual prior events causally necessitate that no prediction occur. Thus, the fact that it is impossible, in this sense, for

actions to be predicted does not conflict with the thesis that actions are determined. Nor is it surprising that the sense of "possible" here under discussion is not important. Using the notion of "causally-compossible-relative-to-all-actual-prior-events" it turns out, assuming determinism, that only actual events are possible. But it is a strange and unduly restrictive notion of "possible" according to which only actual events are possible!

We need, then, a broader notion of possibility, one which allows for non-actual possibles while also taking into account the notion of causal necessity. We can discover a more relevant notion by examining what is often meant in ordinary contexts when we say, counterfactually, "*e* could have occurred." Suppose we say, counterfactually, "The picnic could have been a success." This sort of statement would normally be made with a suppressed "if"-clause. We might mean, for example, "The picnic could have been a success if it had not rained." Now if the only thing which prevented the picnic from being a success was the rain, we are also likely to say, "The picnic *would* have been a success if it had not rained." In the first case we mean that the substitution of non-rain for rain in the course of events would have *allowed* the picnic to be a success; in the second case we mean that this substitution would have *ensured* the success of the picnic. In both cases we are saying that a certain event could have or would have occurred *if* the prior course of the world had differed from its actual course in specified ways.

Although in ordinary contexts we might not pursue the matter further, in order to be systematic we must inquire further: "*Could it not have rained?*" "*Could non-rain have occurred instead of rain?*"

- The actual rain was causally necessitated by actual events prior to the rain. If we are to suppose that it did not rain, we must also make changes (in our imagination) of still earlier events. Carrying this argument to its logical conclusion, it is obvious that whenever a determinist says that a non-actual event *e* "could have" occurred, he must imagine *an entirely new world*. For the picnic to have been a success, it is required that it not have rained. For it not to have rained, the cloud formation would have had to be different. For the cloud formation

to have been different, it is required that the wind velocity (or some other factor) have been different. Etc.

Not only must we change conditions prior to *e*, if we are to suppose *e* occurs, but we probably¹¹ must change events after *e* as well. Had it not rained, a certain other picnic group near us would not have ended their picnic just then. And had they not ended their picnic just then, they would not have left for home just then. And had they not left for home just then, they would not have had an automobile accident when they did.¹² Etc.

The determinist who says, counterfactually, "*e* could have occurred," must construct a whole world to justify his claim. Nevertheless, this gives him a sense of "possible" that allows non-actual possibles. For a determinist, "*e* could have occurred" may be translated as "a causally compossible world can be imagined in which *e* occurs." Normally the determinist will be able to construct worlds resembling the real one to a large extent. But these worlds will never be exactly like our world except for one event only. Any such imagined world will differ from the real world by at least one event for every moment of time. This will be true, at any rate, if the laws governing these imagined worlds are identical with those of the real world. And I shall assume throughout that these laws (whatever they are, exactly) are held constant.

VI

We can now give what I regard as a reasonable formulation of the question: "Is it possible, in principle, to make scientific predictions of voluntary actions?" The formulation is: "Can one construct causally compossible worlds in which scientific predictions are made of voluntary actions?" In saying that this is a "reasonable" formulation of the question, I do not mean that a negative answer to this question would entail that voluntary actions are not determined. I have already pointed out that determinism does not entail that it is physically possible to make scientific predictions of events, including actions. Hence, neither does determinism entail that there are causally compossible worlds in which scientific predictions of

¹¹ I say "probably" because the definition of determinism does not entail that every event is a determinant of some subsequent event. Thus, if not-*e* actually occurred but had no effect on any subsequent event, then we might substitute *e* for not-*e* without changing any subsequent events. However, though determinism does not require it, it is reasonable to assume that every event will have some differential effect on *some* later event or events.

¹² This is all plausible, at any rate, if we deny fatalism. Fatalism, which is by no means implied by determinism, is the view that certain events will happen at certain times *no matter what* antecedent conditions obtain.

actions occur. However, since we are assuming that scientific predictions are physically possible, it would be an important negative result to discover that one cannot construct causally compossible worlds in which scientific predictions are made of voluntary actions. This might not prove that actions are undetermined, but it would suggest a disparity between actions and physical phenomena. For, assuming that scientific predictions are physically possible, it does seem that there are causally compossible worlds in which scientific predictions are made of physical events.

Similar comments are in order on the question, "Can one construct causally compossible worlds in which scientific predictions are made of voluntary actions and in which the agent learns beforehand of the prediction?" Determinism does not entail that there must be such causally compossible worlds. But if no such worlds are constructible—worlds in which "books of life" are found, or things comparable to books of life—one might well claim a disparity between voluntary actions and physical phenomena.

Fortunately, I believe that there *are* causally compossible worlds in which scientific predictions are made of voluntary actions and in which, moreover, the agent learns of (some of these) predictions before he performs the predicted actions. I believe that there are causally compossible worlds in which books of life are written before a man's birth. Inscribed in these books are predictions of the agent's actions, predictions based on laws and antecedent conditions. These predictions are correct even though the agent sometimes reads them before he performs the predicted actions. I shall support my claim that there are such causally compossible worlds by giving a sketch of such a world. Before giving my sketch, however, I wish to examine the structure of prediction-making where the prediction itself has a causal effect on the predicted event. This will be essential in understanding how a "book of life" could be written, even though the writer knows that the agent will read it.

Consider the problem of an election predictor. He may know what the precise results of the upcoming election are going to be, if he makes no public prediction of the election. If he publishes a prediction, however, some of the voters, having found

out what the results will be, may change their votes and thereby falsify his prediction. How, then, can a pollster make a genuinely scientific and accurate prediction of an election? Can he take into account the effect of the prediction itself? Herbert Simon has shown that, under specifiable conditions, a predictor can do this.¹³ Essentially, what the predictor must know is the propensity of the voters in the community to *change* their voting intention in accordance with their expectations of the outcome. If persons are more likely to vote for a candidate when they expect him to win than when they expect him to lose, we have a "bandwagon" effect; if the opposite holds, we have an "underdog" effect.

Let us suppose that a given pollster has ascertained that, two days before the election, 60 percent of the electorate plans to vote for candidate *A* and 40 percent for *B*. He also knows that, unless he publishes a prediction, the percentages will be the same on election day. Further suppose he knows that there is a certain "bandwagon" effect obtaining in the voting community.¹⁴ When the original intention of the electorate is to vote 60 percent for *A*, this bandwagon effect can be expressed by the equation, $V = 60 + .2(P - 50)$, where P is the percentage vote for *A* publicly predicted by a pollster, and V is the actual resultant vote for *A*. Clearly, if the pollster publicly predicts that *A* will receive 60 percent of the vote, his prediction will be falsified. Putting $P = 60$, the equation tells us that $V = 62$. In other words, the effect of the prediction, combined with the original voting intention of the electorate, would result in a 62 percent vote for *A*. However, the pollster can easily calculate a value for P which will make $P = V$. He need only solve the two equations, $P = V$ and $V = 60 + .2(P - 50)$. Such a solution yields $P = 62.5$. Thus, the pollster can publish a prediction saying that 62.5 percent of the electorate will vote for *A*, knowing that his own prediction will bring an additional 2.5 percent of the electorate into the *A* column, and thereby make his prediction come true.

Notice that all the antecedent conditions relevant to the outcome cannot be known until it is known what prediction (if any) the pollster will make. His prediction (or lack of prediction) is itself an im-

¹³ "Bandwagon and Underdog Effects of Election Predictions," reprinted in *Models of Man* (New York, 1957). The requisite condition is that the function relating the actual outcome of the voting to the predicted outcome, given the electorate's original voting intention, be *continuous*.

¹⁴ That this bandwagon effect holds in the community could be discovered either by studying previous elections or by deducing it from "higher-level" generalizations found to be true of the community.

portant antecedent condition. However, one of the crucial determinants of the outcome—viz., the original voting intention of the electorate—is given independently of the pollster's prediction. Thus, while holding that factor constant, the pollster calculates what the outcome of the election *would* be, *if* he were to make certain predictions. By solving the equations given above, he discovers a prediction which, if published, would be confirmed. He thereupon forms an intention to publish that prediction and proceeds to fulfill that intention. Until he forms this intention, he does not know what prediction he will make, and therefore does not know all the requisite antecedent conditions from which to deduce the election outcome. But at the same time he makes the prediction (and perhaps even earlier), he does know all the relevant antecedent conditions and has deduced from these conditions what the results will be. Thus, his prediction of the outcome is a truly scientific prediction.

If someone wishes to predict a single person's behavior and yet let him learn of the prediction, the predictor must employ the same sort of strategy as the pollster. He must take into account what the agent's reaction will be to the prediction. There are several kinds of circumstances in which, having made the appropriate calculations, he will be able to make a correct prediction. (A) The agent learns of the prediction but does not want to falsify it. (B) Upon hearing the prediction, the agent decides to falsify it. But later, when the time of the action approaches, he acquires preponderant reasons for doing what was predicted after all. (C) Having decided to refute the prediction, the agent performs the action conforming with it because he doesn't realize that he is conforming with it. (D) At the time of the action the agent lacks either the ability or the opportunity to do anything but conform with the prediction, though he may have believed that he would be able to falsify it. In any of these four kinds of cases, a predictor would be able to calculate that his prediction, together with numerous other antecedent conditions, would causally necessitate that the agent perform the predicted action. In a case of kind (B), for example, the predictor may be able to foresee that the agent will first read his prediction and decide to falsify it. But other factors will crop up—ones which the agent did not originally count on—which will make him change his mind and perform the predicted action after all. And the predictor also foresees this.

In the first three kinds of cases, (A), (B), and (C),

the agent performs the predicted action *voluntarily* (though in (C) he does not realize that what he is doing falls under the description "what was predicted"). In other words, in each of these three kinds of cases, the agent *could have* acted otherwise, in at least one sense of "could have" which some philosophers think is relevant to free will. Thus, the possibility of a scientific prediction does not require that the agent be *unable* to act in any way different from the prediction. All that is required is that the agent will not *in fact* act in any way different from the prediction. A predictor might know that an agent will in fact act in a certain way, not because he knows the agent will be incapable of doing otherwise, but because he knows that the agent will *choose* or *decide* to act as predicted. This point will be clarified at the end of the paper in a brief discussion of the indicated sense of "could have."

I shall now give a sketch of a causally compossible world in which a large number of correct predictions are made of an agent's behavior. Since I imagine this world to be governed by the same laws as those of the real world, and since I do not know all the laws of the real world, I cannot *prove* that my imagined world really is causally compossible. But as far as I can tell from common-sense knowledge of psychological and physical regularities, it certainly seems to be causally compossible. In this world, predictions of a man's life are made in great detail and inscribed in a "book of life," (parts of) which the agent subsequently reads. Obviously, I cannot describe the whole of this world, but I shall describe some of its most important and problematic features, namely the interaction between the agent and the book. Unfortunately, I shall have to omit a description of another important part of the world, the part in which the predictor (or predictors) gathers his data and makes his calculations. I am unable to describe this part of the world, first, because I do not know all the laws which the predictor would have at his disposal, and secondly, because I am not able to say just what the structure of this being would be. However, the main features of his *modus operandi* should be clear from our discussion of the pollster, whose technique is at the heart of such predicting.

VII

And now to the description of the world.

While browsing around the library one day, I

noticed an old dusty tome, quite large, entitled "Alvin I. Goldman." I take it from the shelf and start reading. In great detail, it describes my life as a little boy. It always gibes with my memory and sometimes even revives my memory of forgotten events. I realize that this purports to be a book of my life and I resolve to test it. Turning to the section with today's date on it, I find the following entry for 2:36 p.m. "He discovers me on the shelf. He takes me down and starts reading me. . . ." I look at the clock and see that it is 3:03. It is quite plausible, I say to myself, that I found the book about half an hour ago. I turn now to the entry for 3:03. It reads: "He is reading me. He is reading me. He is reading me." I continue looking at the book in this place, meanwhile thinking how remarkable the book is. The entry reads: "He continues to look at me, meanwhile thinking how remarkable I am."

I decide to defeat the book by looking at a future entry. I turn to an entry 18 minutes hence. It says: "He is reading this sentence." Aha, I say to myself, all I need do is refrain from reading that sentence 18 minutes from now. I check the clock. To ensure that I won't read that sentence, I close the book. My mind wanders; the book has revived a buried memory and I reminisce about it. I decide to reread the book there and relive the experience. That's safe, I tell myself, because it is an earlier part of the book. I read that passage and become lost in reverie and rekindled emotion. Time passes. Suddenly I start. Oh yes, I intended to refute the book. But what was the time of the listed action?, I ask myself. It was 3:19, wasn't it? But it's 3:21 now, which means I have already refuted the book. Let me check and make sure. I inspect the book at the entry for 3:17. Hmm, that seems to be the wrong place for there it says I'm in a reverie. I skip a couple of pages and suddenly my eyes alight on the sentence: "He is reading this sentence." But it's an entry for 3:21, I notice! So I made a mistake. The action I had intended to refute was to occur at 3:21, not 3:19. I look at the clock, and it is still 3:21. I have not refuted the book after all.

I now turn to the entry for 3:28. It reads, "He is leaving the library, on his way to the President's office." Good heavens, I say to myself, I had completely forgotten about my appointment with the President of the University at 3:30. I suppose I could falsify the book by not going, but it is much more important for me not to be late for that appointment. I'll refute the book some other time!

Since I do have a few minutes, however, I turn back to the entry for 3:22. Sure enough, it says that my reading the 3:28 entry has reminded me about the appointment. Before putting the book back on the shelf, and leaving, I turn to an entry for tomorrow at 3:30 p.m. "He's still riding the bus bound for Chicago," it reads. Well, I say to myself, *that* prediction will be easy to refute. I have absolutely no intention of going to Chicago tomorrow.

Despite my decision to refute the book, events later induce me to change my mind and to conform to it. For although I want to refute the book on this matter, stronger reasons arise for not refuting it. When I get home that evening I find a note from my wife saying that her father (in Chicago) is ill and that she had to take the car and drive to Chicago. I call her there and she explains what has happened. I tell her about the book. Next morning she calls again with news that her father's condition is deteriorating and that I must come to Chicago immediately. As I hang up I realize that the book may turn out right after all, but the situation nevertheless demands that I go to Chicago. I might still refute it by going by plane or train. However, I call the airlines and am told that the fog is delaying all flights. The railroad says that there are no trains for Chicago till later in the day. So, acquiescing, I take a bus to Chicago, and find myself on it at 3:30.

VIII

Let me interrupt my narrative here. I have given several cases in which the book is not refuted, and the reader should be convinced that I could easily continue this way. But it is important now to reply to several objections which the anti-predictionist is anxious to make against my procedure.

(1) "*Your story clearly presupposes determinism. But whether or not determinism is true is the central matter of dispute. Hence, you are begging the question.*" Admittedly, my story does presuppose determinism. Unless determinism were true, the imagined predictor could not have figured out what actions the agent would perform and then written them in the book. However, I do not think that this begs the question. For I am not here trying to prove that determinism is true. I am merely trying to show that the thesis of determinism is quite compatible with the world as we know it and with human nature as we know it. The world

depicted in my story seems to be very much like the real world except that it contains different antecedent conditions. The fact that this imagined world is determined and contains predictions of actions, and yet it resembles the real world very closely, suggests to me that the real world may also be determined. At any rate, this supposition seems quite tenable, and its tenability is what I seek to establish in this paper.

(2) "*The story you told was fixed. Events might have been different from the way you described them. For example, the fog might not have curtailed all air traffic.*" No, events could not be different in the world I am imagining. That is, in my world all the events I described were causally necessitated by prior antecedent conditions. I did not describe all the antecedent conditions, so perhaps the reader cannot see that each event I did describe was causally necessitated by them. But, since it is a deterministic world, that is so. No one can imagine my world and also substitute the negation of one of the events I described. I'm not "fixing" the story by saying that the fog curtailed air traffic; that just is the way my imagined world goes.

(3) "*But I can imagine a world in which some putative predictions of actions are refuted.*" I have no doubt that you can; that is very easy. You could even imagine a world somewhat like the one I have just described, but in which putative predictions are falsified. But this proves nothing at all. I would never deny that one can construct some causally compossible worlds in which putative scientific predictions of actions are not successful. I have only claimed that one can (also) construct some causally compossible worlds in which genuine scientific predictions of actions are made (and are successful). The situation with predictions of action is no different from the one with predictions of physical events. We can construct causally compossible worlds in which predictions of physical phenomena are correct. But we can also construct worlds in which putative scientific predictions of physical phenomena are incorrect. If our ability to construct worlds in which predictions are unsuccessful proves the inherent unpredictableness of the kind of phenomena unsuccessfully predicted, then we can prove the unpredictableness of physical phenomena as easily as the unpredictableness of human action.

(4) "*The world you have described, though possible, is a highly improbable world. Worlds in which putative predictions of actions are falsified are much more probable.*" The notion of one possible world being "more

probable" than another seems to me unintelligible. Surely the statistical sense of probability cannot be intended. There is no way of "sampling" from possible worlds to discover what features most of them have. Perhaps the anti-predictionist means that we can *imagine* more worlds in which putative predictions of actions are falsified. But this too is questionable. I can imagine indefinitely many worlds in which successful predictions of actions are made.

Perhaps the anti-predictionist means that it is improbable that any such sequence of events as I described would occur in the *real* world. He may well be right on this point. However, to talk about what is probable (in the evidential sense) in the real world is just to talk about what has happened, is happening, and will happen *as a matter of fact*. But the dispute between predictionists and anti-predictionists is, presumably, not about what *will* happen, but about what *could* happen *in principle*. This "in principle" goes beyond the particular facts of the actual world.

(5) "*The difference between physical phenomena and action is that predictions of actions can defeat themselves; but predictions of physical events cannot.*" This is not so. One can construct worlds in which the causal effect of a putative prediction of a physical event falsifies that prediction. Jones calculates the position of a speck of dust three inches from his nose and the direction and velocity of wind currents in the room. He then announces his prediction that five seconds thence the speck will be in a certain position. He had neglected to account for the wind expelled from his mouth when he made the prediction, however, and this factor changes the expected position of the speck of dust. Perhaps one can imagine a wider variety of cases in which predictions affect human action more than physical phenomena. But this is only a difference of *degree*, not of kind.

(6) "*Predictions of physical events can refute themselves because the predictor may fail to account for the effect of his own prediction. But were he to take this effect into account, he would make a correct prediction. On the other hand, there are conditions connected with the prediction of action in which, no matter what prediction the predictor makes, his prediction will be falsified. Here there is no question of inaccurate calculation or insufficient information. Whatever he predicts will be incorrect. Yet this situation arises only in connection with human action, not physical events.*"

This is an important objection and warrants detailed discussion.

IX

Suppose that I wish to predict what action you will perform 30 seconds from now, but that I shall not try to change or affect your behavior except by making my prediction. (Thus, I shall not, for example, predict that you will perform no action at all and then make that prediction come true by killing you.) Further suppose that the following conditions obtain. At this moment you want to falsify any prediction that I shall make of your action. Moreover, you will still have this desire 30 seconds from now, and it will be stronger than any conflicting desire you will have at that time. Right now you intend to do action A , but you are prepared to perform \bar{A} (not- A) if I predict that you will perform A . Thirty seconds hence you will have the ability and opportunity to do A and the ability and opportunity to do \bar{A} . Finally, conditions are such that, if I make a prediction in English in your presence, you will understand it, will remember it for 30 seconds, and will be able to tell whether any of your actions will conform to it or not. Given all these conditions, whatever I predict—at least, if I make the prediction by saying it aloud, in your presence, in English, etc.—will be falsified. If I predict you will do A , then you will do \bar{A} , while if I predict that you will do \bar{A} , you will proceed to do A . In other words, in these conditions no prediction of mine is causally compossible with the occurrence of the event I predict. Let C_1, \dots, C_n be the (actual) conditions just delineated, let P_A be my predicting you will do A (announced in the indicated way), and let $P_{\bar{A}}$ be my predicting you will do \bar{A} (announced in the same way). Then *both* sets $\{C_1, \dots, C_n, P_A, A\}$ and $\{C_1, \dots, C_n, P_{\bar{A}}, \bar{A}\}$ are causally *impossible* sets of events.

Notice that this example does not prove that it is causally impossible “simpliciter” for me to make a scientific prediction of your action. All that it proves is that I cannot make such a prediction *in a certain manner*, viz., by announcing it to you in English. The events P_A and $P_{\bar{A}}$ include this particular manner, and that they do so is important. If I predict your action in some other manner, by thinking it to myself or by saying it aloud in Hindustani, for example, the effect on your action would not be the same as if I say it aloud in English. Assume that, if you do not hear me make any prediction or if you hear me say something you fail to understand, you will proceed to perform action A . Then it is causally com-

possible for me to predict your action correctly by announcing the prediction in Hindustani. In other words, letting P_A' be my predicting that you will do A by announcing this in Hindustani, then the set of events $\{C_1, \dots, C_n, P_A', A\}$ is a causally compossible set.

In determining whether or not a certain set of events, including (1) a prediction, (2) the event predicted, and (3) certain other assumed conditions, is a causally compossible set, it is essential to specify the manner of the prediction. This is true *in general*, not just in the case of predictions of action. A prediction which is “embodied” or expressed in one way will not have the same causal effects as the same prediction expressed in another way. We can see this in the case of the speck of dust. Jones predicted the position of the dust by announcing it orally, and this resulted in the falsification of the prediction. But had he made the same prediction in another fashion—say, by moving his toes in a certain conventional pattern—his prediction would not have been falsified, for the position of the dust would not have been affected.

What is the significance of the fact that it is causally impossible, in some circumstances, for a (correct) prediction of an action to be made in a specified manner? First, this unpredictability does not prove that these actions are undetermined. Indeed, the very construction of the case in which no prediction is possible *presupposed* the existence of laws of nature which, together with a given prediction, would result in a certain action. In short, the case under discussion should, if anything, support rather than defeat the thesis that actions are determined. The only reason one might have for thinking the contrary is the assumption—which should by now appear very dubious—that determinism entails predictability. What our present case shows, I think, is that under some circumstances, even a determined event may not be susceptible of being correctly predicted in a specified manner. This fact can be further supported by adducing a similar case connected with purely physical events. And this brings me to my second point: the case produced above does not reflect a peculiarity of human action, since parallel examples can be found among physical phenomena.

Imagine a certain physical apparatus placed in front of a piano keyboard. A bar extends from the apparatus and is positioned above a certain key. (Only white keys will be considered.) If the apparatus is not disturbed, the bar will strike that key at a certain time. Now let us suppose that the

apparatus is sensitive to sound, and, in particular, can discriminate between sounds of varying pitches. If the apparatus picks up a certain sound, the position of the bar will move to the right and proceed to strike the key immediately to the right of the original one (if there is one). Specifically, if the sound has the same pitch as that of the key over which the bar is poised, the bar will move. If the monitored sound has any other pitch, the bar will remain in its position and proceed to strike that key.

Now suppose that someone (or something) wishes to make predictions of the behavior of the apparatus. He wishes to predict what key the bar will strike. But the following restriction is made on the manner in which the prediction is to be made. The prediction must be expressed according to a specific set of conventions or symbols. To predict that the bar will strike middle *C*, for example, the predictor must emit a sound with the pitch of middle *C*. To predict that the bar will strike *D*, he must emit a sound with the pitch of that key, etc. All sound emissions are to be made in the neighborhood of the apparatus. Given this restriction on the manner of prediction, it will be causally impossible for the predictor to make a correct prediction. For suppose that the bar is poised above middle *C*. If he predicts that it will strike middle *C*—that is, if he emits a sound of that pitch—the bar will move and proceed to strike *D*. But if he predicts any other behavior of the bar, for example, that it will strike *D*, the bar will remain in its original position and strike middle *C*.

Admittedly, the manner of prediction I have allowed to the predictor of this physical phenomenon is much more narrowly restricted than the manner of prediction allowed to the predictor of human action. But we could imagine physical apparatuses with a greater degree of complexity, able to "refute" predictions made in any of a wider variety of manners. In any case, the principle of the situation is the same for both physical phenomena and human actions, though the manners of prediction which affect one phenomenon may be different from the manners of prediction which affect the other. The latter difference simply reflects that fact that physical objects and human beings do not respond in precisely the same ways to the same causes. But this is equally true of different kinds of physical objects and of different pairs of human beings.

¹⁵ "An Essential Unpredictability in Human Behavior" in *Scientific Psychology: Principles and Approaches*, ed. by B. B. Wolman and E. Nagel (New York, 1965).

The reader should not suppose that the present discussion in any way vitiates my description of the book of life in Sect. VII. Our present discussion shows that under *some* conditions it is *not* causally compossible to predict a man's action in a way which allows him to learn of the prediction. But there are *other* conditions, such as the ones described in Sect. VII, in which such predictions *are* causally compossible. The existence of the latter conditions suffices to establish the possibility (in principle) of scientific predictions of voluntary actions which the agent hears or reads. Admittedly, it is not always possible to make predictions in this manner. But even when it is impossible to let one's prediction become known to the agent, it does not follow that it is impossible to make the prediction "privately." Thus, suppose you are trying to write a book of my life before I am born. Your calculations might show that if you inscribe certain predictions in the book they will be confirmed. For these calculations might reveal that I shall not read the book, or that I shall perform the actions despite the fact that I shall read the book. If so, you may proceed to write the book, having (scientific) knowledge that it will be correct. On the other hand, your calculations might reveal that, no matter what prediction you inscribe in the book, I shall refute it. In this case, you will be unable to write a book of my life. But you may nevertheless have scientific knowledge of what I shall do! Your calculations may reveal that I shall do a certain sequence of actions, as long as I do not come across any (putative) book of my life. If you decide not to write such a book yourself, and if you know that no one else will, you may conclude (deductively) that I shall perform the indicated sequence of acts.

X

In the previous section we saw that, under certain conditions, it may not be causally compossible to predict a certain action in a specified manner. Recently, however, Michael Scriven has claimed that human behavior exhibits an even more important unpredictability.¹⁶ Scriven writes: "So far we merely demonstrate that human choice behavior can be made at least as unpredictable as any physical system. In an important class of examples . . . , a stronger conclusion is demon-

strable."¹⁶ Scriven's example consists in imagining an agent, *X*, who is contrapredictively motivated relative to a certain predictor, *Y*. That is, *X* wants to defeat any prediction *Y* makes about his actions. Scriven further supposes that *X* knows everything that *Y* knows about him. From this information, *X* figures out what *Y* will predict, or will have predicted, about *X*'s action. In other words, *X* "replicates" *Y*'s prediction; he comes to know what *Y*'s prediction is even though *Y* does not announce his prediction. After figuring out *Y*'s "secret" prediction, *X* proceeds to act otherwise. Scriven concludes: "... the present case is more interesting. The idea that human behavior is 'in principle' predictable is not seriously affected by the recognition that one may not be able to announce the predictions to the subjects with impunity (nor, more generally, can one allow them to be discovered). For one can make the predictions and keep them from the subjects. But in the present case, *one cannot make true predictions at all*. Secret predictions are still predictions; unmakeable ones are not."¹⁷

We must first note that Scriven has given a misleading account of his example in saying that "in the present case, *one cannot make true predictions at all*." (Italics his.) True, a particular person, *Y*, is unable to make correct predictions of *X*'s behavior. But *X*'s behavior is not completely immune to prediction. Scriven's case leaves open the possibility that there is, or was, some other being, *Z*, (who may have lived long before *X*) who predicted *X*'s behavior without *X* knowing of this. In order for *X*'s behavior to be *completely* immune to prediction, *X* would have to know with respect to *every* potential predictor—i.e., everyone who lived during or prior to *X*'s lifetime—what predictions, if any, be made about *X*'s behavior. Anything short of this state of knowledge by *X* would leave open the possibility that some being or other correctly predicted what *X* would do, indeed predicted it scientifically.

Secondly, it is questionable whether Scriven's example shows that human behavior is more unpredictable than physical systems, as he suggests in the passage on page 413. Admittedly, Scriven's case goes beyond my previous example in one respect. In my example only certain manners of prediction lead to the performance of a different

action. In Scriven's example, *any* manner of prediction leads to a different action. This is because even the *minimal* manner of "making" a prediction—i.e., having a future-looking *belief*—is self-defeating. Nevertheless, it may still be possible to duplicate Scriven's human behavior case with physical systems. Suppose, for example, that we found neuro-physiological states that correlated with beliefs. That is, suppose we found one-one correlations between a person's believing certain propositions and his being in certain neuro-physiological states. We might then "hook up" a physical system to a potential predictor in such a way that the state of the system is causally affected by the beliefs (or their neuro-physiological correlates) of the predictor. The physical system would be arranged so that whenever the potential predictor had a belief about a future state of the physical system, this belief would cause the system to go into another state instead.

The third and most important point I wish to make is based on a criticism of Scriven by David K. Lewis and Jane Shelby Richardson.¹⁸ In the competition between Scriven's agent and predictor, each is trying to get sufficient information about the other and to calculate from this data just what the other will do (or believe). Let us combine both of these factors—data and calculation—and call them "knowledge," since the function of the calculations is to add to the calculator's knowledge. Lewis and Richardson argue forcefully that it is impossible for *both* the predictor and the agent to have sufficient, or complete, knowledge of his opponent. We can construct two sorts of cases. We can endow the predictor with complete information about the agent, but this forces us to deny to the agent complete knowledge about the predictor. Or we can endow the agent with complete knowledge about the predictor, but in so doing we must deny complete knowledge about the agent to the predictor. Scriven does the latter, although he is not quite aware of this. In saying that the agent is able to "replicate" the predictor's prediction and then decide to act otherwise, he is in effect saying that there is some aspect of the agent's motivational structure which is *unknown* to the predictor. But if there is some fact relevant to the prediction which the predictor does not know, it is not surprising that he makes

¹⁶ *Ibid.*, p. 413.

¹⁷ *Ibid.*, p. 414.

¹⁸ "Scriven on Human Unpredictability," *Philosophical Studies*, vol. 17 (1966).

an incorrect prediction. He simply is not in a position to make a scientific prediction. This hardly shows that the agent's behavior is inherently immune to scientific prediction. It merely shows that this particular predictor, *T*, does not know enough about *X* to make a scientific prediction.

For this reason Scriven's case is less interesting than the one I presented in the foregoing section. In my example, the predictor is unable to make a correct prediction (in a specified manner) even though he has all relevant information. In Scriven's example, the predictor lacks some relevant information. Scriven's example, then, hardly warrants us in concluding that human behavior is undetermined. The fact that someone with insufficient knowledge is unable to predict an event correctly does not at all suggest that the event is undetermined. Of course, Scriven does not claim that his example shows behavior to be undetermined. I say this merely to remind the anti-predictionist that he can take no comfort from Scriven's case.

XI

I have shown that there are causally compossible worlds in which voluntary actions are scientifically predicted. Let us now see whether there are causally compossible worlds in which a person scientifically predicts one of his *own* actions. I think that there are such worlds and I shall illustrate by continuing the description of the world I was sketching earlier.

Having tested my book of life on a very large number of occasions during many months and failed to refute it, I become convinced that whatever it says is true. I have about as good inductive evidence for this proposition as I do for many another proposition I could be said to know. Finally, I get up enough courage to look at the very end of the book and, as expected, it tells when and how I shall die. Dated five years hence, it describes my committing suicide by jumping off the 86th floor observation deck of the Empire State Building. From a description of the thoughts which

will flash through my mind before jumping, it is clear that the intervening five years will have been terrible. As the result of those experiences, I shall have emotions and desires (and beliefs) which will induce me to jump. Since I trust the book completely, I now conclude that I *shall* commit suicide five years hence. Moreover, I can be said to *know* that I shall commit suicide.

As described so far, we cannot consider my prediction of my suicide a "scientific" prediction. To be a scientific prediction the predicted event must be *deduced* from laws and antecedent conditions, while, as I have described the case, no deduction was involved. However, we might supplement the situation so as to include a deduction. The book may be imagined to list the relevant physical and psychological laws (in a footnote, say) and the relevant conditions which determine my committing suicide (my intention to commit suicide, my proximity to the fence surrounding the observation deck, the absence of guards or other interfering factors, etc.). From these laws and conditions I actually deduce my future action.¹⁹

This example shows, contrary to the view of some authors, that we can have inductive knowledge of our own future actions, knowledge which is not based on having already made a decision or formed an intention to perform the future action. Stuart Hampshire, for example, has recently written, "... I cannot intelligibly justify a claim to certain knowledge of what I shall voluntarily do on a specific occasion by an inductive argument; if I do really know what I shall do, voluntarily, and entirely of my own free will, on a specific occasion, I must know this in virtue of a firm intention to act in a certain way."²⁰ The case outlined, I believe, shows that Hampshire is mistaken. In that case, there is a time at which I do have certain knowledge of what I shall do (at any rate, about as "certain" as one can be with inductive evidence) and yet I have formed no intention nor made any decision to perform that action. At the time I read the book's prediction, I do not intend to commit suicide. But although I do not intend to commit suicide, I fully believe and know that, five

¹⁹ That these conditions will actually obtain is, of course, open to doubt. Moreover, I have not learned of *them* by scientific prediction. I have simply "taken the book's word" that these conditions will obtain; I have not deduced them from other, still earlier, conditions. However, there are no restrictions on the manner in which a predictor comes to know antecedent conditions. One way predictors might learn about antecedent conditions is by using various measuring devices and instruments, the reliability of which is supported by inductive evidence. My book of life may be regarded as such a device, and my inductive evidence supporting its reliability may be as strong as that supporting the reliability of various other devices which scientists commonly use for obtaining knowledge of antecedent conditions.

²⁰ *Freedom of the Individual* (London, 1965), see p. 54.

years later, I shall intend to commit suicide. I firmly believe that, at that later time, I shall feel certain emotions and have certain desires which will induce me to jump off the Empire State Building. At the time of my reading the book I do not feel those things, but I commiserate with my future self, much as I commiserate with and understand another person's desires, beliefs, feelings, intentions, etc. Still, my understanding of these states of mind and of the action in which they will issue is the understanding of a spectator; my knowledge of these states and of my future action is purely inductive. Moreover, this knowledge is of a particular *voluntary* act to be performed at a specified time. Though the suicide will be a "desperate" action, it will in no sense be "coerced" or done unknowingly; it will flow from a firm intention, an intention formed very deliberately. But that intention will not be formed until after I have had certain experiences, experiences which, at the time I am reading the book, I have not yet had.

We can imagine two alternative series of events to occur between my reading the book and my suicide. First, I might *forget* what I have learned from the book, and later decide to commit suicide. Secondly, while never forgetting the prediction, the knowledge of my future suicide may gradually change from more inductive knowledge to knowledge based on intention. In this second alternative, there is never any "moment" of decision. I never pass from a state of complete doubt about committing suicide into a sudden intention of committing suicide. Rather, there is a gradual change, over the five-year period, from mere inductive knowledge that I shall commit suicide to an intention to commit suicide. When I first read the book I am fully prepared to assent to the proposition that I shall commit suicide. But I am saddened by the thought; my heart isn't in it. Later, as a result of various tragic experiences, my *will* acquiesces in the idea. I begin to welcome the thought of suicide, to entertain the thought of committing suicide with pleasure and relief. By the time the appointed time comes around, I am *bent* on suicide. This gradual change in attitude constitutes the difference between the kind of knowledge of my future suicide, the difference between mere inductive knowledge and knowledge based on intention. Hampshire claims that the first kind of knowledge of one's own action is

impossible. The present case, I believe, shows this claim to be mistaken.

Many philosophers seem to be very uncomfortable with the idea of a book of life. They believe that the existence of such books—or of foreknowledge of actions in any form—would deprive us of all the essential characteristics of voluntary behavior: choice, decision, deliberation, etc.²¹ I do not think this fear is warranted, however. I have just shown that even if a person reads what a book of life predicts, and believes this prediction, he can still perform the indicated action voluntarily. Moreover, the existence of predictions which the agent does *not* read leaves ample opportunity for deliberation and decision. An agent may know that a book of his life exists and yet proceed to make decisions and to deliberate as all of us do now. The agent's belief that there is such a book, and his belief that the book's existence implies that his actions are causally necessitated, is compatible with his deliberating whether to do one action or another. Although his future action is causally necessitated, one of the antecedent conditions which necessitate it is his deliberation. Indeed, the prediction in the book of life was made precisely because its writer knew that the agent would deliberate and then decide to do the predicted action. Thus, the book of life can hardly be said to preclude deliberation. Nor does the book of life imply that the agent's deliberation is "for naught," or "irrelevant." On the contrary, his deliberation is a crucial antecedent condition: were he not to deliberate, he probably would not perform the action he eventually does perform. Deliberation and decision are perfectly compatible with the existence of books of life; and they are perfectly compatible with the thesis that they, and the actions in which they issue, are determined.

XII

If actions are determined, there is at least one sense in which an agent "cannot" act other than he does: his actual action is causally necessitated by actual prior events, and hence any other (incompatible) action is causally impossible with actual prior events. It is precisely because his action is causally necessitated that it is amenable to scientific prediction. But it is generally accepted that if an agent does an action *A* voluntarily, he

²¹ One such philosopher is Richard Taylor. See his "Deliberation and Foreknowledge," *American Philosophical Quarterly*, vol. 1 (1964), pp. 73-80. Many others could also be named.

also "can" do otherwise. So far we have only provided a sense of "can" in which he *cannot* do otherwise. We are therefore obliged to identify some *other* sense of "can," or "could" in which an agent "can" do actions he does not in fact do. Such a sense of "can" has long been in the literature, having been defended by Hobbes, Edwards, Hume, Mill, Moore, Nowell-Smith, Stevenson, and others.

Suppose that John does *A* voluntarily. To say that he "could have" done otherwise indicates that he had another alternative open to him, the alternative of doing *Ā*. But what is meant by saying that *Ā* was "open" to John? What is meant, I believe, is that, *if*, contrary to fact, John had wanted or chosen or decided to do *Ā*, then he would have succeeded in doing *Ā*. The "alternative" open to John is not the alternative of doing *Ā* given his (actual) desire or decision to do *A*, but the possibility of doing *Ā* relative to a (non-actual) desire to do *Ā*. Here, as elsewhere, the determinist's sense of "could have" involves the supposition of counterfactual conditions; indeed, if taken to its logical conclusion, a whole counterfactual world. In the "could have" pertaining to action, the main counterfactual feature pertains to the agent's desire, intention, or decision. But the analysis of "could have" is not wholly counterfactual. To say that John "could have" done *Ā* is also to make a

categorical assertion about the real world, viz.; that John *was able* to do *Ā*, that he *had* the ability and the opportunity to do *Ā*. John's actual ability and actual opportunity were such that, if his desire or intention to act had been different, his action would have been different. Thus, the analysis of "John could have done *Ā*" would be formulated as: "John had the ability and opportunity to do *Ā*, and (therefore) if he had decided to do *Ā*, he would have done *Ā*."

A scientific predictor who predicts that someone will perform a voluntary action *A* may thus be justified in accepting the following two propositions: (1) The agent *will* (certainly) do *A*, and (2) The agent *could* do *Ā*. (2) is warranted because, as the predictor realizes, the agent will have the ability and opportunity to do *Ā*. (1) is also warranted, however, because, as the predictor knows, the agent will *in fact* choose to do *A*, not *Ā*. The predictor knows that he will choose to do *A* because he has deduced that he will so choose from antecedent conditions and psychological laws having choices (or desires) as their dependent variables.

The question remains whether the specified sense of "could" is *the* sense relevant to the traditional problems of freedom and responsibility. My own opinion is that this is the relevant sense, but I have no new arguments to give on this score. Here I wish not to join the fray, but to leave it.

The University of Michigan

Received March 22, 1967

II. CRAIGIAN TRANSCRIPTIONISM

C. A. HOOKER

I

IN the present context, I shall use the term "Instrumentalist" to refer to anyone who claims that the specifically *theoretical* terms of scientific theories do not enlarge the ontological commitment of the theory over and above that already implied by the *observational* terms of those theories. Instrumentalists usually claim that, with regard to its specifically theoretical aspect, the theories of science are to be regarded as *computational devices* only.

Obviously the Instrumentalist is free to make his point against the Realist (who, I shall suppose, subscribes to the negation of the Instrumentalist's position) in whatever way he finds appropriate, and "Transcriptionism"¹ is a position which he can adopt for these purposes. The pivotal point of the Transcriptionist's position is his assertion that for each member of some well defined class of scientific theories there can be constructed another, logically distinct, theory such that this constructed theory does not itself contain any theoretical terms but nevertheless retains all of some class of "desirable features" that the original theory possessed. The Instrumentally committed Transcriptionist is then in a position to attempt an Ockhamist argument against the Realist along the following lines:² Theoretical terms are eliminable and the empirical content of scientific theories can be stated in a language³ involving the existence of observable entities and properties only; therefore, there are (logically) no grounds for regarding theoretical terms as empirically meaningful. It is important to note that the Transcriptionist's thesis is a *logical* one, and not merely factual in status.

This fact will be relied upon at various points in the paper and a simple argument for it will be given in Sect. IV. The process of obtaining the constructed theory from the original I call "transcription."

I have spoken of the Transcriptionist as if it were obvious that he is a committed Instrumentalist. That is, indeed, the role in which he is usually cast.⁴ It is also the more interesting, because the more controversial, of the positions open to him. But I must now draw attention to the fact that the Transcriptionist need not regard himself as committed in this manner. Instead, it is open to him to remain *agnostic* on the question of the status of theoretical terms. Thus, he may assert what I have called above the "pivotal point" of the Transcriptionist position, but refrain from going on to develop the Ockhamist argument against the Realist. In the light of the possibility of the systematic replacement of theories by "transcribed variants, he may rather conclude that the question of the status of theoretical terms must be left entirely open. In this way any explicit commitment to a specific thesis regarding the status of theoretical terms is avoided. This makes the position stronger, if less interesting, than the former. In what follows, most of what is said will apply equally to both positions without further comment being necessary. Where further comment is necessary, it will be given explicitly.

It is clear that discussion of the Transcriptionist's position cannot proceed until some definite transcriptional procedure is specified. Only then does it become a definite question what the class of transcribable theories is and what features of these theories are preserved unchanged under transcrip-

¹ I was first introduced to this term by Mr. J. S. Mills, to whom in general I am indebted for the initial impetus to write this paper and for many helpful discussions concerning its content. I am indebted to Mills for emphasizing the possibility of an agnostic Transcriptionist position and the rejecting of inductivist moves outright. Mills also suggested the Argument A_1 and at the same time rejected it for the reasons given. I am also indebted to Dr. H. Lacey for valuable discussion on the more technical aspects of Craig's work. I also wish to thank the referee for the paper for his valuable and detailed comments.

² The argument is to be stated more precisely in Sect. IV below.

³ For convenience, I shall often regard that fragment of a natural language which is used in some scientific theory as itself a language.

⁴ Thus Ernest Nagel, for example, refers to Transcriptionism only in the context of its supporting Instrumentalism. See [8], p. 134ff.

tion. The relevance of Craig's theorem to the Transcriptionist's position is that it offers, *prima facie*, the precise transcriptional procedure which is required. Craig himself [2] was the first to discuss his theorem in detail from this point of view and subsequent discussion has included N. Goodman [3], C. G. Hempel [4, 5], G. Maxwell [7], E. Nagel [8, 9], H. Putman [10], I. Scheffler [11, 12], and J. J. C. Smart [13]. I wish to point out that I shall in no way attempt to call in question the formal status of Craig's work or to detract in any way from its significance for the fields of mathematics and logic. What it is my intention to do in this paper is to discuss the Craigian Transcriptionist position in general and give a critical analysis of two of the arguments which surround the Transcriptionist's use of Craig's theorem. In so doing I shall attempt to evaluate the Transcriptionist's position. The formal content and the proof of the theorem in question is given by Craig in [1] and an expanded, less formal, account is offered by Craig in [2] and by Putnam in [10]. In what follows I shall assume that the reader is familiar with Craig's theorem.

II

There is no doubt that, because of the manner of construction, the Craigian-transcribed, or C-T, theory of any theory has a high degree of complexity and seems to carry with it a certain air of artificiality. In the *first* place, an entire dimension of complexity is introduced because of the method of construction of the C-T theory. The sort of complexity involved here can be seen by observing that if A is a theorem of some theory, then so also is $A \& A \& \dots \& A$. Similarly, $\vee A \vee \dots \vee A$, $\vee A \vee B \vee \dots \vee M$ are also theorems, where $A, B, \dots M$, are sentences of the theory.⁵ In general, if A is a theorem of the theory, then certain large classes of truth functional combinations of A 's and other sentences of the theory will also be theorems of the theory. But each of these theorems will require a slightly different proof and hence each theorem will have a distinct Gödel number assigned to it. Thus, for every such theorem, so long as A and the other sentences involved contain only observational terms, there will be a corresponding axiom in the C-T theory. A *second* dimension of complexity is introduced in the following way. In general, a scientific theory

will contain a number of axioms having at least one universal quantifier and containing only theoretical terms. The theory will also contain a further set of postulates in which is given necessary, or sufficient, or necessary and sufficient conditions for the application of one or more theoretical terms on the basis of certain observable data. But existing scientific theories, unless they are totally general, will also include references to specific individuals and/or conditions. And real individuals not specifically mentioned by the theory may also be introduced as specific instantiations of the theorems of the theory. Now all of the resulting sentences containing these specific terms, being theorems of the original theory, will be included as axioms of the corresponding C-T theory. Suppose, for example, we have a theory T_1 which consists of the postulates $(A_1): (\forall x) (O_1x \supset Tx)$; $(A_2): (\forall x) (Tx \supset O_2x)$; $(A_3): O_1a$; where ' O_1 ' and ' O_2 ' are observational predicates and ' T ' is a theoretical predicate, whilst ' a ' refers to some real individual. Then T_1 has, besides the sentence $(\forall x) (O_1x \supset O_2x)$, the sentences $O_1a \supset O_2a$, O_1a and O_2a , as observational consequences. These latter three will appear in the C-T theory of T_1 along with $(\forall x) (O_1x \supset O_2x)$. Thus the set—possibly denumerably infinite—of all sentences giving true descriptions of all real individuals may be added to the theory to yield, as consequences of the theory, a possibly denumerably infinite set of sentences. (Here "individuals" includes states, processes, and other "initial conditions" as well as concrete objects. Note that this complexity is distinct from the truth functional sort first discussed.)

Now it seems to me that neither of these dimensions of complexity is essential to the Transcriptionist's *philosophical* position, and if they could be removed this would be a considerable improvement in the presentation of the Transcriptionist's position. In what follows I shall first indicate how the second of these may be removed and then discuss the difficulties in removing the first.

Though what I have hitherto called the transcribed *theory* may certainly include all of the individual instantiations of its theorems and their consequences which I have mentioned, it seems to me that a narrower concept of "theory" can be legitimately distinguished. From this narrower point of view it is the business of theories to formulate the *general* features of the world's structure and law-like behavior, and to disclose its

⁵ Here ' \vee ' stands for logical disjunction, ' $\&$ ' for conjunction.

fundamental ontology, i.e., the types of things there are. It is not the business of theories, qua theory, to be concerned with any specific individuals, or individual instances of its general pronouncements. From this point of view theories will contain no sentences which contain expressions which refer to specific individuals. It is a straightforward matter to exclude the introduction to a theory of specific instantiations of its general expressions by individuals which it does not itself mention. However, it might still seem that there are some, less general, theories which unavoidably violate the sense of "theory" here proposed. I refer to those theories which incorporate special circumstances into their formulation. To take an example, suppose that a restricted gravitational theory G for just the solar system is drawn up with the axioms:

- (A_1) Any two massive bodies attract each other with a force varying inversely as the square of the distance between them and directly as the product of their masses.
- (A_2) M (to stand as an abbreviation for Newton's three laws of motion).
- (A_3) The sun and the planets (Mercury, Venus, Earth, Mars, etc.) are gravitating (i.e. massive) bodies.

Here (A_3) contains the names of specific individuals. But we may simply drop (A_3) and retain the completely general theory consisting of (A_1) and (A_2) alone.

In general, individually referring expressions will arise in a theory in one of two ways. They may simply be introduced as the theoretically postulated possessors of some theoretical property T . (Cf. (A_3) above) Such individually referring expressions are removed from the theory simply by dropping these postulates (as above). On the other hand individually referring expressions may turn up in a theory imbedded in the theoretical sentences and not postulationally introduced as above. For example, sentences of the form $(\forall x)(\forall y)(xRy \ \& \ xRa \supset Fy)$ may occur in a theory, where the individually referring term a is "imbedded" in the sentence. (I am thinking especially of biological and sociological theories where paradigm species, epochs, etc., may occur.) In these cases there will always have been evidential grounds on which the individual a was first introduced (at least this will be so in a responsibly constructed theory). If the above sentence is not to be merely an isolated generalization, but part of a coherent theory then a will have been introduced because it has been

verified that a has some property, or properties, ' O_1 ' say, and there is an essentially generalized sentence of the theory which contains the predicate ' O_1 ' and which may be instantiated by a to form the above sentence. (An "essentially generalized" sentence is to be one which makes no reference to specific individuals.. For example, the above sentence may have been derived from the sentences O_1a and $(\forall z)(O_1z \supset (\forall x)(\forall y)(xRy \ \& \ xRz \supset Fy))$. In these cases we replace the sentence containing the imbedded individual by the conjunction of a sentence making no reference to specific individuals and a sentence giving the evidential warrant for introducing the individual in the first place. Then this latter sentence is dropped from the theory, leaving only the essentially generalized sentence.

The point I wish to make is that I am advocating a programme for all theories in which sentences referring to specific individuals are removed from those theories either by simply dropping them or by replacing the original sentence by a universally generalized one together with some (at least one) "correspondence rule" giving the deductive or inductive observational warrant for the drawing of the original theoretical sentence as conclusion. These latter sentences are then simply dropped, for they strictly belong to the evidence which we bring to the theory and not to the theory itself. These procedures remove the individually referring expressions from the statements of a theory, whether or not the individuals referred to are observable. Of course we are concerned, in the present context, primarily with removing expressions uniquely referring to observable individuals so that such expressions do not appear in the C-T theory. This frees the C-T theory of all reference to particular individuals. I shall characterize as *Restricted* theories restricted by the removal of all expressions containing reference to specific individuals, (including states, epochs, processes, etc.).

As soon as the notion of a restricted theory is formulated it is seen to have positive advantages for the formulation of the C-T theory. For the C-T theory of a restricted theory will contain in its axioms no sentences containing terms which refer to specific individuals. Thus the second of the two dimensions of complexity discussed above will have been completely removed. In this case it will, in fact, begin to look more like a C-T theory, for it will now contain, besides logical truths, only conjunctions of essentially general sentences. That is, it will preserve only the general structural connections between observable aspects of the

world, and the general law-like behavior that is observable in that domain.⁶

These simplifying moves may have made the C-T theory a little more palatable as a *theory*. But it is desirable to effect a complete simplification, i.e., that the first or "logical" dimension of complexity be also eliminated. Unfortunately, this proves to be impossible.⁷

Such simplifying moves as come readily to mind are easily shown to break down. For example, any set of rules designed to replace truth-functionally compound sentences in the C-T theory by their constituents must always fail to do so if applied a finite number of times, since the C-T theory will contain truth-functionally compound expressions of any finite order of complexity. Indeed, even were such simplifications possible there would still occur, for every sentence *A*, all sentences which are logically equivalent to *A*. But since there is no effective decision procedure in the predicate calculi for determining logical equivalence, the set of sentences could not be any further simplified. On both grounds, therefore, no effective simplifying procedures exist.

But the failure of simplifying moves does not prevent us from enquiring about the nature of the putative end product of such simplifying moves. If there were to be an effectively completable simplification program, then we should finish with a set of sentences containing all and only those logically distinct sentences which are theorems of the original theory and which contain no theoretical terms. I shall refer to this set as the S-C-T set (the Simplified C-T set). It is, I submit, highly unlikely that any genuine restricted scientific theory is, or will be, such that the *non-logical* part of its S-C-T set contains an infinite number of members. To take the theory T_1 as an example the one and only non-logical member of the S-C-T set of T_1 is $(\forall x) (O_1x \supset O_2x)$. However, I know of no way of *proving* any relevant theorems here even for formally developed systems, let alone for the relatively informal theories of current science.

But if it is in fact very improbable that a genuine and restricted theory of science should violate this condition of finitude, then this is of some importance for the Transcriptionist. For I believe that the theory which the Transcriptionist should

be aiming to construct is just that theory whose axioms and theorems are the members of the S-C-T set. It is an interesting question, both in its own right, but for the Transcriptionist in particular, whether there is an alternative effective procedure for constructing this set, and whether, and under what conditions, the non-logical part of the set is, as I have suggested, finite. In conclusion, it seems possible, intuitively desirable, and I hope to show, useful for the Transcriptionist to formulate his thesis in terms of restricted theories and, at least, the simplest form of the C-T theory effectively available. (I shall call this latter the PS-C-T theory, the Partially Simplified C-T theory.)

III

The extant literature abounds with argued rejections of the usefulness and/or relevance of Craig's theorem for an Instrumentalist position. (I take it that these same arguments would be levelled against the agnostic form of Transcriptionist position as well.) However, it seems to me that upon closer examination a great many of these arguments are a great deal weaker than might be thought and that a re-evaluation of the Transcriptionist position is in order. The arguments can be roughly divided into two groups, those that are predominantly logically based and those that are not. I am of the opinion that the latter group of arguments have no force against the Transcriptionist position whatever. Elsewhere (Hooker [6]), I have argued that they demonstrate at most that there are certain *pragmatic-psychological* disadvantages associated with the Transcriptionist's position. But even if true, such arguments are logically irrelevant to the truth of the Transcriptionist's position which rests upon *logical* and not psychological grounds. However, I am not concerned to argue this further here. In this paper I wish to examine those arguments which seek to demonstrate *logical* defects in the Transcriptionist's position. There are two such arguments. I shall also examine the question of whether or not a putative Transcriptionist success bears *logically* upon the Realist/Instrumentalist dispute. I shall argue that it does not.

⁶ It is also in keeping with the spirit of these simplifications that we demand of the mathematical aspects of a restricted theory that they also remain in uninstantiated form. Roughly, in those cases where functional laws are involved, they remain with no substitution instances involving individual numbers.

⁷ Hempel mentions that no formal method exists for deciding whether the C-T theory will have only a finite number of axioms, [4], n. 60, p. 93 and [5], p. 699.

p2-318

In this section I shall consider the two predominantly logically-based arguments against the Transcriptionist. By way of preparation for these arguments I shall first discuss the notions of the deductive and inductive systemization which a theory effects among the empirical data of its domain. The notion of deductive systemization is clear enough. Consider, for example, the theory T_1 of Sect. II above. This theory yields a deductive systemization among observation sentences involving the predicates ' O_1 ' and ' O_2 ' in virtue of the fact that from O_1a & T_1 the observation sentence O_2a can be derived. Thus O_1a and O_2a are deductively associated via T_1 . In general, then, because of its logical structure, a theory provides systematic, deductively derivable, connections among observation sentences and among observational data in general. The question of the deductive systemization of an original theory vis-a-vis its C-T theory has already been answered: the deductive systemization of the two theories is identical with one another.

Inductive systemization is similarly conceived. It may happen that certain sentences of the theory lend themselves to providing probabilistic, or confirmatory, support to certain other sentences while not yet connecting these sentences deductively. Thus, for example, consider a theory T_2 whose axioms are:

- (A_1) $(\forall x) (Px \supset Gx \& Tx \& Ex \& Vx \& Sx)$;
 (A_2) $(\forall x) (Px \supset Ix)$;
 (A_3) $(\forall x) (Ix \supset Fx)$;

where ' P ' and ' I ' are the only theoretical terms.⁸ Suppose now that the observational evidence $Ga \& Ta \& Ea \& Va \& Sa$ is introduced. Then, under appropriate conditions, we might conclude "Very probably, Pa ." But $Pa \supset Ia \supset Fa$ in T_2 . Therefore, very probably, Fa . That is the linking "If $Ga \& Ta \& Ea \& Va \& Sa$ then very probably Fa " is supported by T_2 and this expresses an inductive systemizing power on the part of T_2 . This particular example can be generalized to yield a general notion of inductive systemization.

Now the Transcriptionist is compelled to admit, if his position is even to be plausible, that preservation of deductive systemization under transcription is required. Otherwise his transcribed theory will

have deleted (or created!) empirical content, and this is unacceptable on any account. But must the Transcriptionist also agree that the preservation of inductive systemization under transcription is required of him?

Not necessarily. He might reject such a demand on the ground that the whole question of inductive procedures is, fundamentally, a *psychologically* determined affair. In this case he can reject the demand to preserve inductive systemization outright, since it is a demand to preserve a *logically and empirically inessential* feature of scientific theories. (Though it might be conceded to be a psychologically important feature of those theories.) He *might* argue in this way. If he did so, he would be in the company of a good many philosophers, so that this maneuver is not at all arbitrary. On the other hand, he *might* wish to accept the possibility of an acceptable, logically based development of inductive procedures. In this case he must agree to accept the requirement in question. Even so, he is entitled to be cautious about statements and arguments based upon the assumption of alleged inductive principles *until* a clearly acceptable, objective, account of inductive procedures has actually been given. But because it will allow the Transcriptionist greater freedom of movement if he can accept the requirement that all intuitively acceptable inductive systemization be preserved under transcription, I propose to continue the discussion on that basis. The two arguments to be examined here both purport to show that the C-T theory does not have the inductive systemizing power which the original theory possessed.

(A_1) *The Argument from Analyticity*

This argument has not been raised, in the literature, specifically against the Craigian Transcriptionist. But both Hempel, [4], p. 80, and Scheffler, [12], pp. 221ff, raise it against the Ramseyan Transcriptionist.⁹ It is a simple matter to re-interpret the argument so as to apply to the Craigian version.

As against the Ramseyan version, the argument runs: Consider the Theory T_2 above and introduce the same observational evidence, viz., $Ga \& Ta \& Ea \& Va \& Sa$. We have already seen that this evidence inductively supports, via T_2 , the con-

⁸ The example is in fact taken from Hempel, [4], pp. 78-79. (Scheffler, [12], p. 220, also uses it). In the original version, the axiom (A_1) was expanded into five distinct axioms. The example was intended to have a prosaic interpretation in terms of tests for phosphorous, ('P').

⁹ It is beyond the scope of this paper to enter into a discussion of this form of Transcriptionism. Both Hempel and Scheffler give introductory discussions and full references.

clusion Fa . But the Ramsey Sentence corresponding to the theory T_2 is:¹⁰ $(S_1): (\exists\phi) (\exists\psi) (\forall x) (\phi x \supset \psi x). (\phi x \supset Gx \& Tx \& Ex \& Vx \& Sx) \& (\psi x \supset Fx)$. S_1 is logically true. (To show this, we choose " ϕx " to be " Fx " and " ψx " to be " $Fx \& Gx \& Tx \& Ex \& Vx \& Sx$ ".) But now the statement: $(S_2): (\exists\psi) (\exists\phi) (x) ((\phi x \supset \psi x) \& (\phi x \supset Gx \& Tx \& Ex \& Vx \& Sx). (\psi x \supset \neg Fx))$, where " $\neg Fx$ " means " x is not F ," is also logically true. (To show this, repeat the proof for S_1 but with " $\neg F$ " substituted for " F ".) Thus if it is claimed that S_1 supports an inductive inference from $Ga \& Ta \& Ea \& Va \& Sa$ to Fa , (in order to make S_1 agree with T_2 so far as inductive systemization is concerned), then it must also be acknowledged that S_2 supports an inductive inference from the same evidence to $\neg Fa$, the negation of the former conclusion. But, from the point of view of the transcribed theories, there is no way to choose between S_1 and S_2 . Therefore, Ramseyan transcription does not preserve inductive systemization.

The application of the argument to the Craigian form of Transcriptionism should now be clear. The theory T_2 has, as it stands, no deductively derivable observational consequences whatsoever. And, regarded as a restricted theory, it can have no inductively derivable observational consequences either. It therefore follows that the C-T theory of T_2 will contain only logical truths and hence will be logically true. But now the same argument applied to S_1 above may now be applied to the C-T theory of T_2 (or its derivatives) to show that, if any inductive conclusions at all are supported, there is no way to choose between contradictory conclusions.

A seemingly simple and obvious rebuttal of this argument consists in pointing out that a criterion for deciding between various transcribed theories is indeed available, viz., that of checking the predicates of the transcribed theory against those appearing in the original theory. Thus while ' F ' occurs in the original theory T_2 above, ' $\neg F$ ' does not occur there. Therefore, this suggestion runs, S_1 above and its C-T theory correlate is to be preferred over S_2 and its C-T theory correlate as preserving the predicates of the original theory. However, this reply is unacceptable for at least two reasons. To begin with the weaker of these, this move seems to grant a status to the original theory which is too "high," as regards content and truth, to be consistent with the Transcriptionist position. For now it is the original theory which decides, in

regard to matters pertaining to empirical fact, which of two statements, themselves of the same logical status, is to be accepted. Under these conditions it seems to follow that the original theory retains some empirical content not captured by the C-T theory. But there is a much stronger reason for rejecting this move. Recall that the C-T theory will contain *only* logical truths and it will contain *every* logical truth derivable from the axioms of the original theory. But for every logically true statement containing the predicate ' F ' (say) and other predicates there will be a corresponding logically true statement containing the predicate ' $\neg F$ ', with all other predicates remaining unchanged. Thus, in the above example there will not be *two* C-T theories, but only *one*, and hence there are no choices to be made. And we can now develop the argument so that it concludes: Therefore, if the C-T theory supports any inductive inferences at all, it supports, *simultaneously*, contradictory inferences.

The Transcriptionist is not totally without reply to the argument. But it will be best for him to admit at the outset that the C-T theory of T_2 , being logically true, does not provide support, deductive or inductive, for anything. It has no empirical content whatsoever. Now this admission may seem to work in the Realist's favor until it is realized that it also reflects back upon the status of the original theory T_2 . Thus T_2 has of itself, as we have noted, neither deductively nor inductively derivable observational consequences. Its *sole* function is to *suggest* a possible ordering among observable data. But it then follows that, in a very good sense, T_2 has no empirical content, and without empirical content can it be claimed that it is an *acceptable ground* for systematizing anything? After all, the predicates ' P ' and ' T ' are not tied down in any convincing way and any number of inductive connections might be arbitrarily established in this manner. But now if we introduce whatever evidence it was which led to the construction of T_2 in the first place the only generalized statement which it will *directly* support is something like: "With high probability and for all x if $Gx \& Tx \& Ex \& Vx \& Sx$ is true then so also is Fx ." But if this is all the "theory" amounts to, it can be easily accommodated by the Transcriptionist. Finally, the Transcriptionist can point out that if it is only those theories with no empirical content (in the above sense) which his transcription procedure

¹⁰ The Ramsey Sentence is the transcribed theory of this form of Transcriptionism.

fails to deal with adequately, then this is not a very serious difficulty for him. For even the Realist will find these theories of little interest and they can be set aside as an eccentric and insignificant aspect of the whole.

Certainly, so long as the theory has some non-trivial, i.e., non-logically true, observational consequences, its corresponding C-T theory will be non-logically true and the argument will not apply. When this latter condition is not fulfilled the above argument has force only if a sharp distinction is drawn between deductive and inductive empirical content, i.e., it must be claimed that a theory may possess a *genuine* inductive systematizing power for empirical data even though it possesses no deductive systematizing power whatsoever for empirical data. But this distinction can be denied. That it is not entirely implausible to deny the distinction will be shown during the consideration of the next argument.

(A₂) *The Argument from the Elimination of Postulates Under Transcription.*

Scheffler develops, (see [12] p. 215), for the purposes of this argument, a theory T_3 which has the following postulates:

- (P₁): $(\forall x) (\forall y) (\forall z) ((Bxyz \& Mz \& Rz) \supset (Mx \& My))$;
 (P₂): $(\forall x) (Mx \supset (\forall y) (Fxy \supset Cxy))$;
 (P₃) $Bbca \& Ma \& Ra$;

where 'M' is the only theoretical term and 'a' & 'b' and 'c' stand for observable individuals. Clearly T_3 has observational consequences, viz.:

- (C₁): $(\forall y) (Fby \supset Cby)$, and
 (C₂): $(\forall y) (Fcy \supset Ccy)$, and
 (C₃): $Bbca \& Ra$.

The observational evidence,

- (E₁): $(Fae \& Cae) \& (Faf \& Caf) \dots (Fam \& Cam)$
 (E₂): $Bbca \& Ra$,

is now introduced in the case of (E₁), or explicitly distinguished within the theory in the case of (E₂).

Scheffler argues that (E₁) inductively supports the statement $(\forall y) (Fay \supset Cay)$ and that this statement in turn inductively supports the statement Ma (from (P₂)). But (E₂) conjoined to Ma is (P₃) which, we have seen, yields the observational consequences (C₁) and (C₂), via (P₁) and (P₂). Thus in the presence of (T_3), (E₁) & (E₂) inductively

support (C₁) and (C₂). But, Scheffler argues, the C-T theory of T_3 contains only logical truths and conjunctions of the statements (C₁), (C₂), and (C₃). In particular, it contains none of (P₁), (P₂), or (P₃). But we have seen that the establishment of the above inductive connection relies upon two uses of (P₂) and one of (P₁). It follows, Scheffler argues, that the C-T theory cannot support these inductive connections since it does not contain two crucial links in the chain used to establish them.

Again, Scheffler develops a variant, T'_3 , of T_3 in which the postulates are (P₁), (P₂), (E₁), and (E₂) above. Again the C-T theory of T'_3 does not include either (P₁) or (P₂), [it consists only of conjunctions of (E₁) and (E₂) and logical truths], and hence it cannot be used to effect the inductive link which, Scheffler claims, holds between (E₁) & (E₂) and (C₁) and (C₂). In general, since no matter what observational evidence is added to (P₁) and (P₂) the corresponding C-T theory will not contain either (P₁) or (P₂), the C-T theory will not support any inductive links established upon the basis of (P₁) and (P₂).

There is, however, an immediate move which can be made against Scheffler's example as it stands and the argument based on it. Neither of the theories T_3 , T'_3 are restricted theories, for both contain postulates mentioning specific individuals. But I have already argued that the Transcriptionist may reasonably confine his thesis to restricted theories. If the offending postulates are removed from the respective theories, the resulting restricted theories coincide; in fact they both reduce to just the postulates (P₁) and (P₂). Let us call this latter theory RT_3 . But RT_3 has no deductively or inductively derivable observational consequences of a non-trivial kind. It is thus open to rejection on just the same grounds as the theory T_2 above was rejected.

However, the present argument, unlike that of (A₁) above, does not rely in any way upon the Transcribed theory being logically true. Thus it should be a relatively simple matter to "correct" the example so that these "inessential difficulties" are avoided. Now the only theoretical term of RT_3 is 'M'. Thus the only way in which to give RT_3 non-trivial, deductively derivable, observational consequences is to add a further postulate to the theory of the form:

$$(P_4): (\forall x) (Ox \supset Mx),$$

which gives an observationally sufficient condition, ("O—"), for the applicability of the theoretical

predicate 'M'. Then this new theory, T_4 say, certainly does have deductively derivable empirical content, for example:

$$(C_4): (\forall x) (Ox \supset (\forall y) (Fxy \supset Cxy)).$$

Now suppose that 'O' is chosen so that it does not establish deductive connections between (E_1) and the other observational predicates of the theory. It is still the case, however, that T_4 supports the same inductive link between $(E_1) \& (E_2)$ and $(C_1) \& (C_2)$. For (E_1) still inductively supports Ma and this, together with (E_2) , yields (C_1) and (C_2) , via (P_1) and (P_2) .

What now of the C-T theory of T_4 ? As before, it must fail to contain either (P_1) or (P_2) . Can it then be used to support the same inductive connections as T_4 ? The answer is that it can! For now the PS-C-T set contains, among other sentences, the following:

$$(C_5): (\forall x) (\forall y) (\forall z) ((Oz \& Bxyz \& Rx) \supset (u) (Fxu \supset Cxu));$$

$$(C_6): (\forall x) (\forall y) (\forall z) ((Oz \& Bxyz \& Rx) \supset (u) (Fyu \supset Cyu)).$$

Now in the original argument, (E_1) was taken as supporting the statement $(\forall y) (Fay \supset Cay)$. For consistency, this result must be kept here. And in the original argument the statement $(\forall y) (Fay \supset Cay)$ stood as the consequent of an instantiation of a universally quantified material implicational statement and, as such, it was held to support inductively the antecedent of the consequent for that particular instantiation. This roughly describes the *weakest* form of the inductive rule which Scheffler appears to be using—although he does not discuss the question of what rules are to be accepted in the present context. I shall call it R . Now the correct conditions for the application of R , using $(\forall y) (Fay \supset Cay)$, are realized under the instantiation of (C_4) for the individual a . Therefore, (E_1) inductively supports Oa . But $Oa \& E_2$ entails both (C_1) and (C_2) via the instantiation of (C_5) and (C_6) for the individual a . Scheffler's argument for an inductive link between $(E_1) \& (E_2)$ and (C_1) and (C_2) can be exactly paralleled in the C-T theory and its derivatives. Thus, even though neither P_1 nor P_2 survive in the C-T theory, the particular sort of inductive systemization which Scheffler considers is preserved under transcription.

¹¹ From the Transcriptionist point of view, the avoidance of Scheffler's argument here should in itself constitute good grounds for defending this position.

¹² I have not stated in what formal calculus the statements corresponding to the implications are to be formulated. Typically the predicate calculus would be chosen and then, corresponding to $O \rightarrow T$ we would have: $(\forall x) (Ox \supset Tx)$.

It would definitely seem to be the case therefore that Scheffler's argument here rests crucially upon one or both of the following; (i) The use of only those theories having no non-trivial observational consequences, (ii) the introduction of purely existential statements mentioning specific individuals as postulates of the theory. The reasonableness of rejecting the second of these features has already been argued. And the very fact that the status of the C-T theory changes so sharply as we pass from those theories without, to those theories with, non-trivial observational consequences is itself good grounds for regarding the former theories as in some important sense defective to "second class" theories. This fact provides the additional backing promised for the argued rejection of these defective theories under argument (A_1) above.¹¹

However, while this latter point can be given some additional weight by pointing out that what was illustrated in the above example remains true, under certain conditions, for all theories and all inductive rules of the type R , it must be admitted that the conditions which must obtain for this to be true may perhaps be regarded as a little too stringent to be acceptable. The basic principles operating here can be illustrated as follows. If a restricted theory is to have non-trivial observational consequences then there must be given at least one observational, sufficient condition for the applicability of one or more theoretical predicates. Typically, there will occur in the theory inference chains of the following types:

$$\begin{array}{l} (Ch_1) \quad O_1 \rightarrow T_1 \begin{cases} \rightarrow T_2 \rightarrow T_3 \rightarrow \dots \rightarrow O_2 \\ \rightarrow T_4 \rightarrow T_5 \rightarrow \dots \rightarrow O_3 \end{cases} \\ (Ch_2): \quad O_1 \rightarrow \begin{matrix} T_1 \\ O_2 \end{matrix} \begin{cases} \rightarrow T_2 \rightarrow T_3 \rightarrow \dots \rightarrow O_3 \\ \rightarrow T_4 \rightarrow T_5 \rightarrow \dots \rightarrow O_4 \end{cases} \end{array}$$

where ' \rightarrow ' symbolizes implication and the ' O_i ' and ' T_i ' are observational and theoretical predicates respectively.¹² In general, one would expect complex, interlocking patterns consisting of these basic chains. But these simple chains will be sufficient to illustrate the issues adequately.

In the first case, (Ch_1) ; the PS-C-T set will have the sentences corresponding to $O_1 \rightarrow O_2$ and $O_1 \rightarrow O_3$ in it, whilst in the case of (Ch_2) there will

occur $O_1 \rightarrow O_3$ and $O_1 \rightarrow (O_2 \rightarrow O_4)$. Thus the inference chains are "compressed" under transcription. Assuming universally quantified statements corresponding to the implication links and/or a suitable extension of the inductive rule R, we observe that inductive support "flows backwards" up the implicational arrows. Thus we have, under R or its extension:

(Ch'_1): $C_3 \rightarrow \dots \rightarrow T_5 \rightarrow T_4 \rightarrow T_1 \rightarrow T_2 \rightarrow \dots \rightarrow O_2$ and therefore $O_3 \rightarrow O_2$. Similarly, $O_2 \rightarrow O_3$.

(Ch'_2): $C_4 \rightarrow \dots \rightarrow T_5 \rightarrow T_4$, therefore $O_4 \& O_2 \rightarrow T_1 \rightarrow T_2 \rightarrow T_3 \rightarrow \dots \rightarrow O_3$ and therefore $O_4 \& C_2 \rightarrow O_3$. Similarly, $O_3 \& O_2 \rightarrow O_4$.

Here ' \rightarrow ' symbolizes inductive support. But corresponding to the case (Ch'_1) we have, in the C-T theory, $O_3 \rightarrow O_1 \rightarrow O_2$ and therefore $O_3 \rightarrow O_2$. Similarly, $O_2 \rightarrow O_1 \rightarrow O_3$ and so $O_2 \rightarrow O_3$. And corresponding to the second case, (Ch'_2), we have $O_2 \rightarrow (O_1 \rightarrow O_4)$. Therefore if O_2 then $O_4 \rightarrow O_1 \rightarrow O_3$. Therefore, if O_2 then $O_4 \rightarrow O_3$ and so $O_2 \& O_4 \rightarrow O_3$. Similarly we also have $O_3 \& O_2 \rightarrow O_4$. Thus compression of implication chains preserves inductive links of the type considered between the extremes of the chains.

But the examples (Ch_1) and (Ch_2) are restricted in one crucial sense. They both commence and terminate with observational predicates. Consider the following example:

(Ch_3): $T_1 \rightarrow T_2 \rightarrow T_3 \rightarrow T_4 \rightarrow \dots \rightarrow O_2$
 \searrow
 O_1

This example is similar, as it stands, to the restricted form of the theory T_3 above in that it too has no observational consequences, but is claimed to support the inductive relation $O_2 \rightarrow O_1$. We obtain a satisfactory version of (Ch_3) by replacing it by:

(Ch_4): $O_3 \rightarrow T_1 \rightarrow T_2 \rightarrow T_3 \rightarrow T_4 \rightarrow \dots \rightarrow O_2$
 \searrow
 O_1

Now all inductive links are preserved when the theory in which (Ch_4) occurs is transcribed. But suppose that instead of (Ch_4) we introduce:

(Ch_5): $T_1 \rightarrow T_2 \rightarrow T_3 \rightarrow T_4 \rightarrow \dots \rightarrow O_2$
 \searrow
 O_1 $O_3 \nearrow$

Any theory which contains a chain corresponding to (Ch_5) certainly has observational consequences, for example, the statement corresponding to $O_3 \rightarrow O_2$. Thus it is satisfactory from this point of view. And yet it is clear that the transcribed theory of that theory will contain nothing to support an

inductive link between O_2 and O_1 . (Though it will preserve the link $O_2 \rightarrow O_3$.) But *both* of these links are supported by an original theory containing (Ch_5). In order to avoid this difficulty, the Transcriptionist will have to further restrict the types of theories which he is prepared to accept. The restriction will run somewhat as follows.

(Res.): A theory is acceptable only if it contains no implication chain such that there is at least one observationally *necessary* condition for the applicability of a theoretical term farther to the left than the last observationally *sufficient* condition for the applicability of a theoretical term, all terms belonging to the chain in question.

In general, we feel that an observationally sufficient condition for the applicability of a theoretical term ties it more closely to reality than does an observationally necessary condition for its applicability. If there are terms in a theory which have only observationally necessary conditions for their application then we may "inflate" the theory in a more or less arbitrary manner whilst preserving all deductive and inductive links amongst observation sentences. For example, the simple theory consisting only of the sentence $(\forall x)(T_1x \supset O_1x)$ may be "inflated" to $(\forall x)(T_3x \supset T_4x)$ and $(\forall x)(T_4x \supset O_1x)$ and $(\forall x)(T_1x \supset T_3x)$. There is, as it were, no observational checks on the freedom of theory construction in this case. Whereas, in the case of those theories all of whose terms have observationally sufficient conditions for their applicability, there are in general very much more stringent restrictions placed upon acceptable extensions of the theory. For this very reason, some may feel that the above requirement is too stringent whilst others may feel that it is only reasonable. A Realist is likely to reject this intuitive claim altogether as placing too much emphasis upon the *observable* world, whilst his opponent is likely to find it more acceptable. Thus despite the fact that it is unlikely that agreement can be reached regarding this restriction, it does not seem to be a totally implausible position to defend. I conclude therefore that the force of Scheffler's argument here has been considerably weakened.

Of course, there may be other, more sophisticated, types of inductive inference which the process of Craigian transcription fails to preserve; but now the onus is on the opponent to demonstrate this. And it is always open to the Transcriptionist to demand a formal statement of the

logical machinery which governs these inductive principles so that he can incorporate it into theories and test whether or not inductive links are indeed preserved under transcription. It seems reasonable, our intuitions here being as unreliable as they are, that the Transcriptionist refuse to concede defeat until failure to preserve acceptable inductive links has been formally and objectively shown.

It should also be added that there is another feature which strengthens the Transcriptionist's position here. All theories are built on evidence and a great many of the inductive moves which we wish to make are dictated by the statistical features of the evidence. Even if it then happens that the form of the theory which is built upon that evidence reflects the desired inductive moves, it will also often be true that they could have been formulated on the basis of the evidence alone.¹³ But every inductive rule formulated solely in the observational language will appear in the transcribed theory and thus none of these is lost. I conclude therefore, that the *arbitrary* supplementation of the transcribed theory by inductive rules, (Scheffler, [12], p. 202), in order to save it from the general argument (A_2) which we have been considering, a move which Scheffler rightly rejects, is found to be unnecessary (at least in part). Moreover, qua process of supplementation on the basis of what the evidence suggests, the comments made immediately above show that it need not be at all arbitrary.

This completes my examination of the two logically-based arguments against the Craigian Transcriptionist, neither of which, it seems to me, is ultimately compelling. At least this is so if the Transcriptionist is sufficiently careful in the formulation of his position. In summary, the Transcriptionist's thesis is that all and only restricted theories which have ordered observational support (i.e., obey "Res.") and at least one non-trivial observational consequence, are transcribable, and that both deductive and inductive systemization is preserved under transcription for these theories.

IV

In this section I shall consider the question of whether the putative success of the Transcrip-

tionist position as stated immediately above has any bearing upon the conflict between the Realist and the Transcriptionist (of either variety). One might suppose (or hope), for example, that a successful defence of the Transcriptionist position entails the untenability of the Realist's position. In this connection I want first to consider some arguments against Craigian Transcriptionism which Putnam, [10], raises.

Putnam dismisses Craigian Transcriptionism as irrelevant to the Realist's position on the grounds that, in theorizing, our aim is not merely to systematize, but to talk truthfully about what there really is, including centrally theoretical entities.¹⁴ But the Transcriptionist prevents us from doing this. Therefore, the Transcriptionist position is to be rejected. Putnam admits that this rebuttal of the Transcriptionist with what he calls the "short answer" might be argued to be *too short* on either of the following grounds: (i) That the *existence of theoretical entities* is thereby assumed or presupposed, and (ii) that the *meaningfulness of theoretical terms* is thereby assumed or presupposed.

Objection (i) is dismissed on the following grounds. So long as the Transcriptionist merely contends for the *contingent falsity* of the Realist's position, then there is overwhelming evidence to support the view that this contention is itself false, viz., all the evidence pointing to the, admittedly possible, existence of theoretical entities. Putnam's move against objection (ii) is simply to deny the Instrumentally inclined Transcriptionist's assertion that theoretical terms are meaningless. He does this on the ground that many of these terms have been accepted into our common language and this acceptance and free use of a term is paradigmatic of what it is for a term to have meaning.¹⁵

Putnam's reply to objection (i) is, I think, correct. It demonstrates the necessity for formulating the Transcriptionist thesis as against the Realist as a *logical*, and not an empirical, thesis. But Putnam's reply to objection (ii) is too short—at least, too short to convince anyone except those deeply committed to upholding some form of (common-language) linguistic philosophy. The argument works, if, *but only if*, the opponent will accept the premiss that whatever terms occur in common language are, necessarily, meaningful when taken in the sense intended by the speakers of that

¹³ The inductive move based upon the theory T_2 above, for its suggested prosaic interpretation in terms of tests for phosphorous, is surely of this kind.

¹⁴ Cf. here the contrasting attitude of Nagel [8], p. 137, and that of Goodman [3], p. 318.

¹⁵ Putnam argues, for example, that terms such as "God" have meaning on just these grounds.

language.¹⁶ But why is this premiss not open to denial? Indeed, why should anyone even attempt to defend its *truth*, let alone its necessity? Is it really logically impossible for common language to contain strictly meaningless terms? Commonly used religious and ethical language, for example, is notorious for the contradictions to which it gives rise. Is every term commonly employed in religious and ethical discourse meaningful as originally intended? The history of philosophy is rife with accusations of nonsense against commonly held views—were *all* of these accusations *logically* false? (It is beside the point to weaken the thesis so that it claims merely that if a term occurs in a common language then it is probably meaningful. In the first place the Transcriptionist's assertion, that certain terms are in fact without meaning, is compatible with this claim. But, more relevantly, the Transcriptionist is making his point on *logical* grounds, so that factual considerations of which terms do and do not, as a matter of fact, have meaning are irrelevant here. Nor is it the case that a position such as Putnam's is essential to the Realist's position.)

The Transcriptionist raises a logical argument for his position and against the Realist's position. What is required of an opponent is that he meet this argument with another argument which either leads to the denial of one of the premisses, or to a denial of the validity, of the former argument. What Putnam has done is to deny the *conclusion* of the Transcriptionist's argument without in any way indicating any error in the Transcriptionist's own argument. On the other hand, it does seem to me that Putnam's general conclusion, viz., that the success of the Transcriptionist's *formal* program (only) is irrelevant to the real core of the dispute between Realists and Transcriptionists, is correct. I shall now attempt to set out an explicit argument to this conclusion, commencing with the Instrumentally committed Transcriptionist.

By way of preparation, the Ockhamist argument employed by the Transcriptionist needs to be given explicit formulation. There will doubtless always be controversy over how strong this principle can be allowed to be whilst still remaining acceptable.

I suggest the following formulation as probably acceptable to all:

(Oc.): For any domain to which we have at least some access, entities within that domain which are such that there is no evidential support whatsoever for their existence, ought, logically, not to be introduced to our theories of that domain.

Now suppose that the Transcriptionist develop his argument as follows:

(P₁): Theoretical terms are, logically, always satisfactorily eliminable from every theory of the class of transcribable theories.

(P₂): All genuine scientific theories can be reasonably developed in a manner which is consistent with their belonging to the class of transcribable theories.

(P₃): Oc.

(P₄): Therefore, no theoretical term ought, logically, to be introduced into scientific theories.

Let it be granted that (P₁), (P₂), and (P₃) are acceptable to Realist and Transcriptionist alike. Further, let it be granted that to accept (P₄) is to accept the Instrumentalistic Transcriptionist's position and to deny the Realist's position.¹⁷ Does this now mean that the Transcriptionist is victorious? *It does not*, for the simple reason that the argument as it stands is *invalid*, it is in fact a *non sequitur*. The additional premiss required for validity is:

(P₅): Entities with terms corresponding to them which are eliminable under Craigian transcriptions are such that there is no evidential support whatsoever for their existence.

Without (P₅) the argument is invalid. On the other hand, without agreement on the part of both Realists and Transcriptionists as to what are acceptable premisses of the argument the conclusion will not be decisive in establishing the Transcriptionist's position. But the real situation, it seems to me, is that while those premisses on which agreement might or can be obtained, viz.,

¹⁶ We must insist on the condition that the meaning must be as originally intended by the speakers of the language. For if this condition is not retained it allows any amount of semantic reinterpretation of the language to go on whilst retaining the *letter* of the language intact. In particular, the Transcriptionist can speak of theoretical expressions as elliptical formulations of observational expressions. Then Putnam's argument would not establish what he intended it to establish.

¹⁷ Strictly, the Transcriptionist requires a formulation of Oc. that will take him from a denial, on logical grounds, of reasons for introducing a theoretical entity to a denial of the meaningfulness of the theoretical term corresponding to that entity. But for the sake of brevity I omit this additional complication.

(P_1), (P_2), and (P_3), are not sufficient to render the argument valid, the sufficient condition of the argument's being valid includes the introduction of a premiss upon which agreement cannot now be, and never will be, obtained. In particular, whilst the Transcriptionist will accept (P_5), the Realist will reject it.

The Instrumentally committed Transcriptionist will accept (P_5) because he *already believes* that theoretical terms do not have empirical content simply because the only domain in which such content resides is the observational domain. On the other hand, the Realist will reject (P_5), arguing that all that Craig's theorem shows is that, for any theory, a subsection of the theory can always be isolated if desired. But this is of no importance at all, he will argue, since both theoretical and observational domains have equal empirical status. Indeed, we might just as well have "eliminated" observational terms, for all the ontological significance this purely formal maneuver has. Thus the Realist asserts that the Transcriptionist has confusedly taken a merely epistemic disparity for a logical disparity. And the Transcriptionist replies that the Realist fails to discern a logical feature of things which only reflects itself in weakened form at the epistemic level. Obviously the position taken up by each party here is governed by attitudes and beliefs more basic than the premisses of the argument in question. Under these circumstances the formal success of the Transcriptionist will fail to convince the Realist of error.

But the Transcriptionist need not be an Instru-

mentalist; he may be agnostic about the status of theoretical terms. Does this alter his position vis-à-vis the Realist if the formal part of his program is successful? Certainly it *appears* more reasonable to insist that, if the Transcriptionist can make good his formal claims, then one and all, Realists included, should adopt an agnostic stance over the question of the status of theoretical terms. But it is a simple matter to show that this is a false impression and that the situation is not in fact changed from what it was above by this switch of positions. Under this new position the Transcriptionist's argument, *if it is to be valid*, must be from (P_1), (P_2) and a further premiss, (P_6) say, to (P_4), where (P_6) is something like:

(P_6): Entities corresponding to terms which are eliminable under Craigian transcriptions are such that we have no good reason to be anything but agnostic about their existence (and about the status of the corresponding terms).

But now there will be precisely the same disagreement about the acceptability of (P_6) as there was about (P_5) and the formal success of the Transcriptionist will once again not tell against the Realist.

In summary, a carefully formulated Transcriptionist position is reasonably defensible in the face of the extant criticisms, but this fact is of no help in settling the Realist/Instrumentalist dispute in favor of the Instrumentalist, or alternatively, of settling the Realist/Agnostic dispute in favor of the Agnostic.

University of Sydney

Received May 11, 1967

REFERENCES

- [1] Craig, W.: "On Axiomatizability Within a System," *The Journal of Symbolic Logic*, vol. 18 (1953), pp. 30-32.
- [2] Craig, W.: "Replacement of Auxiliary Expressions," *The Philosophical Review*, vol. 65 (1956), pp. 38-55.
- [3] Goodman, N.: "Review of W. Craig," [2] *The Journal of Symbolic Logic*, vol. 22 (1957), pp. 317-318.
- [4] Hempel, C. G.: "The Theoretician's Dilemma," in *Minnesota Studies in the Philosophy of Science*, vol. 2, H. Feigl, M. Scriven, and G. Maxwell (eds). (Minnesota, 1958), pp. 37-98.
- [5] Hempel, C. G.: "Implications of Carnap's Work for the Philosophy of Science" in *The Library of Living Philosophers*, Rudolf Carnap. P. A. Schilpp (ed). (La Salle, Illinois), pp. 685-709.
- [6] Hooker, C. A.: "Five Arguments Against Craigian Transcriptionism," forthcoming in the *Australasian Journal of Philosophy* for 1968.
- [7] Maxwell, G.: "The Ontological Status of Theoretical Entities," in *Minnesota Studies in the Philosophy of Science*, vol. 3 H. Feigl and G. Maxwell (eds). (Minnesota, 1962), pp. 3-27.
- [8] Nagel, E.: *The Structure of Science* (London, 1961).
- [9] Nagel, E.: "A Rejoinder to Putnam," *The Journal of Philosophy*, vol. 62 (1965), pp. 429-432.
- [10] Putnam, H.: "Craig's Theorem," *The Journal of Philosophy*, vol. 62 (1965), pp. 251-260.
- [11] Scheffler, I.: "Prospects of a Modest Empiricism," *The Review of Metaphysics*, vol. 10 (1957), pp. 383-411 and 602-625.
- [12] Scheffler, I.: *The Anatomy of Enquiry* (London, 1964).
- [13] Smart, J. J. C.: *Philosophy and Scientific Realism* (London, 1963).

III. KNOWLEDGE, INFERENCE, AND EXPLANATION

GILBERT HARMAN*

I. INTRODUCTION

THIS paper examines applications of an *empiricist analysis of knowledge*. Without attempting to defend the analysis, I shall assume that it is roughly correct and shall draw some consequences. I shall argue in particular that it suggests solutions of problems in inductive logic and statistical explanation. These applications support the analysis; but I shall also show that the analysis is not completely adequate, since it does not provide for a "social aspect" of knowledge.

I take an analysis to be any interesting set of necessary and sufficient conditions. Although I shall not offer an analysis of the meaning of "know" (whatever that would be), I shall appeal to your intuitions about hypothetical cases. I shall claim, for example, that a person can come to know something when he is told it, or when he reads it in the newspaper. Although I may seem to appeal to what one would ordinarily say about such cases and for this reason may seem to be doing "linguistic analysis," I am interested in what is true about such cases and not just in what we say about such cases. But, since I want to test the assumption that ordinary judgments about knowledge are usually correct, trust your natural inclinations about the cases I describe. Consider what you would naturally and ordinarily judge, if you were not doing philosophy. Fine distinctions made in ordinary judgments become blurred when these judgments are made in a philosophical context.

A rough statement of the empiricist analysis is that all knowledge is based on inference from data given in immediate experience. My strategy is to suppose the rough statement roughly true, to assume that ordinary judgments about knowledge are, on the whole, correct, and to see what sort of theory this leads to. I depart from the empiricist tradition in (at least) one important respect. I take the analysis as a rough statement of what it is to *come to know*. I do not want to say anything in this

paper about so-called "memory knowledge." For simplicity, furthermore, I shall consider only cases in which a person comes to know something when he comes to believe it. In other words, I shall disregard cases in which a person comes to know something he previously believed for the wrong reasons.

There are many relevant things I cannot discuss. For example, I shall not discuss the objection that there is no such thing as immediate experience. (For the purposes of this paper, fortunately, it may not be very important whether the objection is right.) Another objection is that a person's knowledge cannot be based on inferences he is not aware he makes. This deserves detailed consideration, especially since it has not received the same amount of critical attention as the first objection. But in this paper I must limit myself to some rather brief remarks.

II. HOW BELIEF IS BASED ON INFERENCE

In this paper I often use the expression "based on inference," and similar expressions. I do not say that, strictly speaking, the knower actually reasons (although I say this when I am speaking loosely). I say rather that, strictly speaking, his belief is based on reasoning. What a belief is based on depends upon how the belief came about; but belief can be based on reasoning even if the belief is not the result of conscious reasoning.

Consider how people talk about computers. Computers are said to add, multiply, compute, reason, and make use of data, even though no one means by this that some person literally does these things. When we talk about computers, we use words like "reasoning," "inference," "data," etc., in a wider sense than when we talk about people. I suggest that empiricists use the wider sense of these terms when they describe knowledge as based on reasoning from data in immediate experience. Thus psychologists have come more and more to explain human behavior by thinking

* I have discussed the subject of this paper with a great many people. I am especially grateful to Paul Benacerraf, John Earman, Richard Jeffrey, and Saul Kripke. Earman suggested several of the examples. The form of the argument is my own, as is the responsibility for errors.

of people as if they were in part computers. They speak of psychological mechanisms and psychological models. Many psychologists have said that the first step in any good psychological explanation is a description of a mechanism that can duplicate the behavior to be explained.¹ If we think of a person (or his brain) as a mechanism like a computer, then we can ascribe inference and reasoning to that person, in the sense in which computers infer and reason. The conscious inferences a person makes are in the extended sense of the term only some of the inferences he makes. We can in this way make sense of the notion that (loosely speaking) a person is not always aware of the inferences he makes.

Psychological explanation typically describes a mechanism by means of a program or flow chart rather than its physiological realization. The same automaton can be constructed in various ways, with either tubes or transistors for example. Two computers, made of different materials but programmed in the same way, may be said to be in the same state when they carry out the same part of the program. Putnam and Fodor have persuasively argued that psychological states are more like being at a particular place in the program than like having something or other happening in your transistors.²

Suppose that a psychologist wants to describe a mechanism to account for belief formation. Having a particular belief must correspond to the machine's being in a particular state. For example, belief might correspond to the state in which the sentence believed is stored in a certain part of computer memory. The psychologist must propose an hypothesis about how the mechanism comes to be in various states of belief. He must explain how the computer comes to store particular sentences in its memory.

My empiricist claims that belief is the result of reasoning in the sense in which computers reason. The process by which the mechanism comes to store a sentence in its memory is like a reasoning process. Moreover, if none of its belief states correspond to beliefs that it is going through such a process, the computer will not be "aware" that it is going through this process. This represents unconscious inference. Conscious reasoning is repre-

sented in the mechanism when reasoning produces in memory sentences that describe the reasoning process.

Notice that the computer analogy does not provide a method for determining what reasoning belief is based on. I explain in Sects. IV and VI below how to discover such reasoning by appeal to intuitive judgments about when a person knows something. Ultimate confirmation of this approach awaits the development of an adequate psychological model. Part of the argument of the present paper is that appeal to intuitive judgments about knowledge and to the empiricist analysis of knowledge can help in the construction of such a model. (Cf. the final paragraph of Sect. III below.)

I now want to describe two principles that an empiricist must accept if he is to offer a plausible account of knowledge. The first is that all inductive inference infers the truth of an explanation. The second is the condition that the lemmas be true. I shall begin with a brief account of each of these principles.

III. EXPLANATIONS AND LEMMAS

The first principle is illustrated whenever a person infers from certain evidence to an explanation of that evidence. The detective infers that the butler did it, since that's the only way to explain the fingerprints on the gun. A scientist infers something about unit charges in order to account for the behavior of oil drops in an experiment he has done. Since the reasoner must infer that one explanation is better than competing explanations, I say he makes an *inference to the best explanation*. In my view, all inductive inference takes this form. Even when a person infers a generalization of the evidence, his inference is good only to the extent that the generalization offers (or is entailed by) a better explanation of the evidence than competing hypotheses. (But note, I do not say that the explanation must be the best of *alternative* explanations; I say rather that it must be the best of *competing* explanations.)

The connection between explanation and induction is implicit in recent work in inductive logic and the theory of explanation. Goodman has shown

¹ E.g., J. A. Deutsch, *The Structural Basis of Behavior* (Chicago, University of Chicago Press, 1960).

² Hilary Putnam, "Minds and Machines" in Sidney Hook (ed.), *Dimensions of Mind* (New York, New York University Press, 1960); "Robots: Machines or Artificially Created Life?" *The Journal of Philosophy*, vol. 61 (1964), pp. 668-691. Jerry A. Fodor, "Explanations in Psychology" in Max Black (ed.), *Philosophy in America* (New York, Ithaca, Cornell University Press, 1965).

that one can ordinarily infer a generalization of the evidence only if the generalization is lawlike, and Hempel and Oppenheim have pointed out that only lawlike generalizations can explain their instances.³ This provides confirmation of the claim that all inductive inference is inference to the best explanation. More confirmation will be provided later.

The second principle an empiricist must accept, the condition that the lemmas be true, says that a person cannot come to know something by inferring it from something false. In Keith Lehrer's example,⁴ suppose Mary has strong evidence that Mr. Nogot, who is in her office, owns a Ford; but suppose that Mr. Nogot does not in fact own a Ford. Perhaps someone else in her office, Mr. Havit, does own a Ford. Still, Mary cannot come to *know* that someone in her office owns a Ford by inferring this from the false premiss that Mr. Nogot, who is in her office, owns a Ford.

I speak of "lemmas" because the relevant propositions need not be included in Mary's initial premiss. Her initial premisses may be that she has seen Mr. Nogot driving a new Ford, that she has heard him say he owns a Ford, etc., where all of these initial premisses are true. It is false that Mr. Nogot owns a Ford; but that is not one of her initial premisses. It is, rather, a provisional conclusion reached on the way to the final conclusion. Such a provisional conclusion, that is a premiss for later steps of the argument, is a lemma. The condition that the lemmas be true says that, if Mary is to know something by virtue of an inference on which her belief is based, every premiss and lemma of that inference must be true.

Mary's belief will often be based on several inferences, only one of which needs to satisfy the condition that the lemmas be true. For example, she might also possess evidence that Mr. Havit owns a Ford and infer from that that someone in her office owns a Ford. That one of her inferences fails to satisfy the condition that the lemmas be true does not prevent Mary from obtaining knowledge from her other inference. Furthermore, even when Mary explicitly reasons in one particular way, we may want to say her belief is also based on other unexpressed reasoning. If Mary has evidence that Mr. Havit owns a Ford, we may also

want to ascribe the second of the above inferences to her even though she consciously formulated only the first. Sect. VI, below, describes how the inferences we shall want to ascribe to a person will depend upon our intuitive judgments about when people know things.

So, inferential knowledge requires two things: inference to the best explanation and the condition that the lemmas be true. I shall now illustrate and support these requirements with some examples.⁵

I shall describe two cases, the *testimony case* and the *lottery case*, and appeal to your natural non-philosophical judgments about these cases. In the testimony case a person comes to know something when he is told about it by an eyewitness or when he reads about it in the newspaper. In the lottery case, a person fails to come to know he will lose a fair lottery, even though he reasons as follows: "Since there are N tickets, the probability of losing is $(N - 1)/N$. This probability is very close to one. Therefore, I shall lose the lottery." A person can know in the testimony case but not in the lottery case, or so we would ordinarily and naturally judge. In the lottery case a person cannot know he will lose no matter how probable this is. The contrast between the two cases may seem paradoxical, since witnesses are sometimes mistaken and newspapers often print things that are false. For some N , the likelihood that a person will lose the lottery is higher than the likelihood that the witness has told the truth or that the newspaper is right. Our ordinary, natural judgments thus seem almost contradictory. How could a person know in the testimony case but not in the lottery case?

At this point many philosophers would reject one of the ordinary judgments no matter how natural the judgment may be. But such rejection would be premature. My strategy is to ask how beliefs are based on reasoning in the two cases. The only relevant reasoning in the lottery case seems to be deductive. From the premiss that the lottery is fair and that there are N tickets, it follows that the probability of any ticket being a loser is $(N - 1)/N$. One can only deduce the probability statement. No deductive inference permits one to detach the probability qualification from the statement that the ticket will lose. I claim moreover that there is no inductive way to detach this

³ Nelson Goodman, *Fact, Fiction, and Forecast* (Cambridge, Mass., Harvard University Press, 1955). C. G. Hempel and Paul Oppenheim, "Studies in the Logic of Explanation," *Philosophy of Science*, vol. 15 (1948), pp. 135-175.

⁴ Keith Lehrer, "Knowledge, Truth, and Evidence," *Analysis*, vol. 25 (1965), pp. 168-175.

⁵ See also Gilbert H. Harman, "The Inference to the Best Explanation," *The Philosophical Review*, vol. 74 (1965), pp. 88-95.

qualification, since inductive inference must take the form of inference to the best explanation and no explanation is involved in the lottery case.

The testimony case is different. No obvious deductive inference leads to a probabilistic conclusion in this case; and acceptance of the testimony can be based on two consecutive inferences to the best explanation. To see this, consider how we would ordinarily explain our evidence, the testimony. First, we would infer that the speaker so testifies because he believes what he says (and not because he has something to gain by so testifying, or because he has gotten confused and has said the opposite of what he means, etc.). Second, we would infer that he believes as he does because in fact he witnessed what he described (and not because he suffered an hallucination, or because his memory deceives him, etc.).

There is, then, an important divergence between the two cases. In the testimony case, the relevant conclusion can be reached by inference to the best explanation. This is not true in the lottery case. It is the appeal to explanation, over and above any appeal to probability, that is important when a person comes to know a nonprobabilistic conclusion.

A person who believes testimony rarely is conscious of reasoning as I have suggested. But, in the ordinary case, such reasoning must be warranted. For suppose that the hearer had good reason to doubt that the speaker has said what he believes, so that the hearer would not be warranted in reasoning in the required way. Then, even if he accepted what the speaker has said and the speaker has spoken truly, the hearer could not be said to know this. The hearer would also fail to gain knowledge if he had good reason to doubt that the speaker's belief is the result of what the speaker witnessed, since again the hearer could not reason in the required way. My analysis of the testimony case would explain why this reason must be warranted if the hearer is to come to know the truth of what he hears. According to that analysis, the hearer's belief is based on the suggested reasoning; and if his belief is to be knowledge, reasoning must be warranted. Therefore, that the this reasoning must be warranted provides some confirmation of my analysis of the testimony case.

Stronger confirmation arises from an application of the condition that the lemmas be true. Suppose that a person who has no reason not to believe a witness does believe him. The hearer cannot

thereby come to know unless *in fact* the testimony was an expression of what the witness believes and unless *in fact* the witness's belief was the result of what he witnessed. If the witness were to say the opposite of what he believes, a listener could not come to know, even if the witness inadvertently spoke the truth. Nor could he come to know if the witness said what is true as a result of remembering the wrong occasion. The witness's knowledge requires the truth of two explanatory claims. We can understand this if we assume that knowledge in the testimony case is based on the reasoning I have already mentioned and if we apply the condition that the lemmas be true. The two explanatory claims appear as lemmas in that reasoning. These lemmas must be true if the hearer is to gain knowledge from the testimony. The empiricist analysis thus permits us to explain things we might not otherwise be able to explain.

We have, then, a rough analysis of knowledge that involves two principles. If we take the analysis as a working hypothesis, we can apply the two principles in order to learn something about knowledge, inference, explanation, and perception. The discussion of the lottery case *versus* the testimony case has provided one example of such an application. I shall now describe other examples.

Notice that to take the analysis as a working hypothesis in this way is to render it immune to a certain sort of counterexample. According to the analysis, knowledge is based on inference to the best explanation; but in order to determine when belief is based on inference and in order to discover what constitutes good inference to the best explanation, one must appeal to the analysis plus intuitions about when people know things. Therefore, the test of the resulting theory cannot be whether or not it conflicts with one's intuitions about when people know things. (This is only partially correct; see the final section of this paper.) Instead, the theory must be judged by whether it can be developed without appeal to *ad hoc* assumptions in a way that sheds light on epistemological and psychological subjects and whether it does this better than competing alternatives. The next three sections of this paper are meant to suggest some of the range and power of this theory.

IV. APPLICATION TO INDUCTIVE LOGIC

We can use the analysis in finding criteria of good inductive inference. Instead of asking directly

whether a particular inference is warranted, we can ask whether a person could come to know by virtue of that inference. If we identify what can be known with what can be inferred, we can discover something important about "detachment" in inductive logic. A principle of detachment would let us "detach" the probability qualification from our conclusion. If there were no rule of detachment, induction would never permit anything more than probabilistic conclusions. But, as inductive logicians have found, it is difficult to state a rule of detachment that does not lead to inconsistency.

Suppose, for example, that detachment were permitted whenever the evidence made a conclusion highly probable. Thus suppose that we could detach a probability qualification whenever our conclusion had a probability (on our total evidence) of at least $(N - 1)/N$. Since any ticket in a fair lottery among N tickets has a probability of $(N - 1)/N$ of being a loser, the suggested principle of detachment would permit us to conclude for each ticket that it will lose. But we also know that one of these tickets will win, so use of high probability to warrant detachment had led us to inconsistency. Some logicians take this result to show that there should be no principle of detachment in inductive logic.⁶

We can avoid this extreme position if we identify the possibility of detachment with the possibility of knowing a nonprobabilistic conclusion. The testimony case tells us that induction sometimes allows nonprobabilistic conclusions, since in that case a person comes to know such a conclusion. The lottery case shows that the inference to such a conclusion is not determined by the high probability one's premisses give his conclusion, since in the lottery case one can only come to know a probability statement. Detachment is possible in the testimony case but not in the lottery case. I have argued that explanation marks the difference between these cases. In the testimony case a person infers the truth of certain explanations. Not so in the lottery case. The problem of detachment arises through failure to notice the role of explanation in inductive inference. Such inference is not just a matter of probability; one must infer the truth of an explanation. Detachment can and must be justified by inference to the best explanation.

This is not to say that probability, or degree of

confirmation, is irrelevant to inductive inference. We can, in fact, use the empiricist analysis again to discover how induction involves probability. Suppose that John and Sam have tossed a fair coin to determine who will have a new hundred-dollar bill. The new hundreds are easily recognizable, being pink, an innovation of the Treasury Department. An hour later, Peter, who knows about the toss, sees John with a new hundred-dollar bill. Peter realizes that John could have received such a bill in only two ways, the most likely being that he won the toss with Sam. There is also an extremely unlikely way, hardly even worth considering. That morning, as a result of a *Consumer Digest* promotion scheme, some person, chosen at random from the population of the United States, has received the only other pink hundred now in general circulation. The odds are two-hundred million to one that John did not receive the *Digest's* bill. So Peter infers that John won the toss with Sam. He infers that the explanation of John's having the bill is that he won the toss and not that he received the *Digest's* bill. If the explanation is right, an ordinary, natural judgment about the coin toss case would be that Peter knows John won the toss.

If this is correct, it suggests one way in which probability can serve as a guide to the best of several competing explanations. Other things equal, the best one will be the most probable one. If it is sufficiently more probable than the others, then a person may infer the truth of that explanation. If *Consumer Digest* had sent pink hundred-dollar bills to every third person, randomly selected, then Peter could not know John has won the coin toss, since that explanation of John's having the bill would no longer be sufficiently more probable than a competing hypothesis. An important issue is how much more probable the one hypothesis must be if it is to provide knowledge. This question may be pursued by further application of the empiricist analysis; but I shall not do so. I shall instead turn to a different aspect of inductive inference.

A complication must be added to what has been said. The best explanation is more than just a highly probable explanation. It must also make what is to be explained considerably more probable than would the denial of that explanation. That is,

⁶ Cf. Henry E. Kyburg, "Probability, Rationality, and a Rule of Detachment" in Brouwer *et. al.* (eds.), *Proceedings of the 1964 Congress on Logic, Methodology, and the Philosophy of Science* (Amsterdam, North Holland Publishing Co., 1965), and references therein. I shall not discuss Kyburg's own solution, since he retains inductive detachment at the expense of deduction. For him one cannot in general infer deductive consequences of what one accepts.

a weak maximum likelihood principle must be satisfied.⁷

To see this, consider the following case. Terry has received a special certificate if he has won a fair lottery among 1000 people. If Terry hasn't won, then George has given him a duplicate of the winning certificate, since George wants Terry to have such a certificate no matter what. Arthur, knowing all this, sees Terry with a certificate. Why cannot Arthur infer that George gave Terry the certificate? That explanation of Terry's having the certificate is very probable; but Arthur cannot make such an inference, because he cannot come to know by virtue of that inference that Terry didn't win the lottery. The most probable explanation does not make what is to be explained any more probable than the denial of that explanation does. That George has given Terry the certificate would make it certain that Terry has a certificate; but this is just as certain if George has not given it to him, because Terry has then won the lottery. Since Terry would have a certificate in any event, Arthur cannot infer that it came from George, even though this explanation is the most probable.

So two things are necessary if an explanation is to be inferable. First, it must be much more probable on the evidence than its denial. Second, it must make what is to be explained more probable than its denial does. This amounts to a synthesis of two apparently conflicting approaches to statistical inference. The Bayesian approach is reflected in the requirement that the best explanation be more probable on the evidence than its denial. The maximum likelihood approach is reflected in the requirement that the explanation make what is to be explained more probable than its denial does.⁸

More needs to be said about this since even these two conditions are not sufficient; but further investigation would place us in the middle of the theory of confirmation. Enough has been said to show how the analysis may be used to study induction from an unusual point of view.

V. APPLICATION TO THE THEORY OF EXPLANATION

If we exploit the connection between explanations and projectible (or inferable) hypotheses, we

may use the analysis to study explanation. An hypothesis is directly confirmed by evidence only if it explains the evidence. So, an hypothesis is a potential explanation if it is the sort of thing that can be directly inferred; and the legitimacy of an inference can again be determined by the possibility of obtaining knowledge by virtue of that inference.

One can show, for example, that a conjunction does not always explain its conjuncts. Let one conjunct be that this is a ticket in a fair lottery among N tickets. Let the other conjunct be that this ticket loses. It is easy to show that the conjunction (that this is a ticket in a fair lottery among N tickets and will lose) cannot explain its first conjunct (that this is a ticket in a fair lottery among N tickets). The result is perfectly obvious, of course, but I want to show how to use the empiricist analysis to demonstrate such a result.

The argument is simple. If the conjunction provides an explanation, then it sometimes provides the best explanation. But then we ought to be able to know something we cannot know. We ought to be able to know in the lottery case that we have a losing ticket; and we cannot know this. If the conjunction provided the best explanation of our evidence, a person in the lottery case could infer the truth of the conjunction from this evidence. In that way he could come to know that his ticket will lose. Since he can't come to know this, the conjunction does not explain its conjunct.

To prove that the conjunction, if an explanation, sometimes satisfies the requirements on the best explanation, notice that it always satisfies the first requirement. The evidence makes the conjunction more probable than not, since the conjunction has a probability on the evidence of $(N-1)/N$. Furthermore, there will be situations in which the weak maximum likelihood principle is satisfied. Typically, in fact, the falsity of the conjunction would make it very improbable that this is a ticket in an N ticket lottery. So, if the conjunction can explain, it can be the best explanation.

This result is trivial and obvious, but the same method can be applied to less trivial cases. It is especially useful in the study of statistical explanation. Consider, for example, the most basic question, whether there can be such a thing as statistical explanation at all. Use of the empiricist

⁷ An explanation of the maximum likelihood principle with further references appears in Ian Hacking, *Logic of Statistical Inference* (Cambridge, Cambridge University Press, 1965).

⁸ The Bayesian position is forcefully presented in Richard Jeffrey, *The Logic of Decision* (New York, McGraw Hill, 1965). The maximum likelihood principle is defended against the Bayesians in Hacking, *op. cit.*

analysis shows there can be and also shows what sort of explanation it is.

Consider cases in which a person comes to know something by means of statistical sampling methods. Suppose, for example, that there are two batches of widgets such that about 70 percent of the widgets in one batch are defective and only about 1 percent of the widgets in the other batch are defective. Confronted with one of the batches, David must decide whether it is the largely defective batch or the good batch. He randomly selects ten widgets from the batch and discovers that seven out of the ten are defective. He infers correctly that this is the defective batch. In this way he comes to know that this is the defective batch, or so we would naturally judge. To apply the empiricist analysis requires assuming his inference is to the best explanation; and to assume this is to assume that there can be statistical explanation. David must choose between two explanations of the makeup of his sample. Both are statistical. Each explains the sample as the result of a random selection from among the items of one of the batches. The explanation David accepts is much more probable than its denial, given the sample he has drawn and assuming that before he had the sample either batch was equally likely. The same explanation makes David's having drawn such a sample more likely than this is made by the explanation he rejects. Therefore, the explanation he accepts is the best explanation of his evidence, and he can come to know the truth of that explanation. He could not, on the empiricist analysis, make his inference if there were no such thing as statistical explanation.

This kind of statistical explanation does not always make what it explains very probable. It is possible, given David's evidence, that the explanation of the makeup of his sample is that he drew randomly from the good batch and this was one of those times when the unlikely thing happens. Such a possibility contradicts the Hempelian account of statistical explanation,⁹ so I shall elaborate.

I can make my point clearer if I change the example. Suppose Sidney selects one of two similar looking coins, a fair one and a weighted one such that the probability of getting heads on a toss of the fair coin is $1/2$ and the probability of getting heads on a toss of the weighted coin is $9/10$. To discover which coin he has, Sidney tosses it ten

times. The coin comes up heads three times and tails seven times. Sidney correctly concludes the coin must be the fair one. We would ordinarily think that Sidney could in this way come to know he has the fair coin. On the empiricist analysis, this means he has inferred the best explanation of that distribution of heads and tails. But the explanation, that these were random tosses of a fair coin, does not make it probable that the coin comes up heads three times and tails seven times. The probability of this happening with a fair coin is considerably less than $1/2$. If we want to accept the empiricist analysis, we must agree that statistical explanation sometimes makes what is to be explained less probable than its denial. This means one has not explained why three heads have come up rather than some other number of heads. The explanation is of a different sort. One explains, as it were, how it happened that three heads came up, what led to this happening. One does not explain why this happened rather than something else, since the same thing could have led to something else.

Suppose Stuart walks into the casino and sees the roulette wheel stop at red fifty times in a row. The explanation of this may be that the wheel is fixed. It may also be that the wheel is fair and this is one of those times when fifty reds are going to come up. Given a fair wheel one expects that to happen sometime (although not very often). But, if the explanation is that the wheel is fair and this is just one of those times, it says what the sequence of reds is the result of, the "outcome" of. It does not say why fifty reds in a row occurred this time rather than some other time, nor why that particular series occurred rather than any of the $2^{50}-1$ other possible series.

I am inclined to suppose that this is the only sort of statistical explanation. But that is another story. I do not want to pursue the theory of explanation in detail. My point has been that the empiricist analysis can be used in the study of explanation and that it results in conclusions different from those generally accepted.

VI. DISCOVERING INFERENCES BELIEF IS BASED ON

Another way to use the analysis exploits the condition that the lemmas be true in order to discover what reasoning knowledge is based on. I begin with a simple example. Normally, if a hearer

⁹ C. G. Hempel, "Aspects of Scientific Explanation" in his *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science* (New York, The Free Press, 1965), esp. pp. 376-412.

is to gain knowledge of what a witness reports, the witness must say what he does because he believes it; and he must believe as he does because of what he saw. Two conditions must thus be satisfied if the hearer is to know. If we wanted to discover the hearer's reasoning, we could use the fact that there are these conditions. We could explain these conditions if we were to assume that they represent lemmas in the hearer's reasoning, since that would make the conditions special cases of the condition that the lemmas be true. Thus we can often account for conditions on knowledge, if we assume that the knowledge is based on relevant reasoning and if we apply the condition that the lemmas be true.

One example worth pursuing, although I shall not say much about it, is knowledge one gets from reading the newspaper. Suppose a misprint changes a false statement into a true one (by, perhaps, substituting the word "not" for the word "now"). In any ordinary case, one cannot come to know by reading that sentence even though the sentence is true. Our method tells us to assume that this fact about misprints represents a lemma in our inference. And it does seem reasonable to assume we infer that the sentence we read is there as a result of the printer correctly forming the sentence that appears in the manuscript. What else do we infer? We ordinarily do not make detailed assumptions about how the reporter got his story, nor about whether the story comes from wire services or from the paper's own reporters. If we are to discover just what we do infer, we must make extensive use of the condition that the lemmas be true. We must discover what has to be true about the way the story gets from reporters to the printer and what has to be true about the way the reporter got his story. We must then associate these conditions with the condition that the lemmas be true, in order to discover what we infer when we come to know by reading the paper. But I shall say nothing more about this problem.

Now consider a case of perceptual knowledge in which a person, as we say, *just sees* that something is true. It is obvious that there are conditions to be satisfied if a case of seeing is to be a case of seeing that something is true. We can account for some of these conditions if we assume that direct perceptual knowledge is based on reasoning. Suppose that Gregory sees a table in the room. As many philosophers have noted, ordinarily, if he is to see *that* there is a table in the room, it must look to him as if there is a table in the room. Further-

more, there must be some causal relationship between the table and its looking to Gregory as if there is a table in the room. It will not do if there is a mirror between Gregory and the table such that he is really seeing the reflection of a different table in a different room. Nor could Gregory see that there is a table if he was hallucinating, even if, by some coincidence he hallucinated a scene exactly like the one in fact before him.

Applying the analysis, we assume that such direct perceptual knowledge is based on inference and attempt to apply the condition that the lemmas be true. This leads us to say that perceptual knowledge is based on inference from data in immediate experience, where such data include how things look, sound, feel, smell, taste, etc. The relevant reasoning infers an explanation of some aspect of immediate experience. In the example, Gregory reasons that it looks as if there is a table because there is a table there and he is looking at it. If he is to reach the conclusion that there is a table, he needs the explanatory statement as a lemma. That is why the truth of the explanatory statement is required if Gregory is to see that there is a table in the room. A similar analysis applies to other cases of direct perceptual knowledge.

I have been purposefully vague about immediate experience, because the empiricist analysis can probably be adapted to any conception. It can apply even if one denies there is any such thing as immediate experience, for one can speak about stimulations of sense organs instead. If Gregory is to see that there is a table in the room, then his eye must be stimulated in a way that depends in part on the table in the room. I can imagine an empiricist who holds that perceptual knowledge is based on inference from immediate stimulation.

Two things must always be remembered. First, an empiricist analysis is not necessarily an analysis of meaning. It is merely an interesting set of necessary and sufficient conditions. It is irrelevant to an empiricist analysis whether the meaning of knowledge claims implies anything about stimulation of sense organs. Second, knowledge can be based on reasoning even when no one actually reasons.

Usually the relevant reasoning will be reasoning only in the sense in which computers reason. The computer analogy is particularly useful if perceptual knowledge is analyzed in terms of stimulations rather than immediate experience, since stimulations are data only in the sense in which a computer can be supplied with data. One might

think here of a computer used to aim antiaircraft missiles in the light of data obtained by radar.

VII. KNOWLEDGE OF THE EXTERNAL WORLD?

Philosophers have wanted to avoid this conception of perceptual knowledge, because they have thought it leads to scepticism. If a person has only his immediate experience to go on, how can he know there is a world of objects surrounding him? How does he know it is not a dream? How does he know it is not the creation of an evil demon?

The problem, if there is one, is not just how one comes to know there is a world of objects, for it arises in any instance of perceptual knowledge. I can see that there is a table in the room only if I can infer an explanation of my immediate experience. How can I legitimately make this inference? How can I rule out the possibility that I may be dreaming? How do I know that a demon psychologist has not attached my brain to a computer that stimulates me as if I were seeing a table? If veridical perception is to provide the best explanation of my experience, that explanation must be more probable than the others. But how can I assume that it is more probable without begging the question? How can I know I have not had many dreams just like this? How can I know I have not had many experiments played on me by the demon psychologist?

Notice that we have no independent way to discover the likelihoods of the various explanations. If one applies the empiricist method for dealing with problems in inductive logic, he may take the fact of perceptual knowledge to show that the hypothesis of veridical perception is highly probable on a person's evidence. The empiricist can in this way avoid the problem of our knowledge of the external world, indeed he can exploit the problem for his own ends in order to argue that there is a predilection for veridical perception built into our confirmation function.

I have tried to show how the empiricist analysis can be used to study induction and explanation and to account for certain requirements on knowledge as special cases of the condition that the lemmas be true. I have described how the analysis can lead one to say that even direct perceptual knowledge is based on inference. In my opinion, the applications of the empiricist analysis show that there must be something to that analysis. I shall

now show that the analysis does not provide the whole story and that it leaves out a "social aspect" of knowledge.

VIII. THE "SOCIAL ASPECT" OF KNOWLEDGE

An empiricist assumes that whether a person knows depends only on the data that person has and not on the data someone else has. There are qualifications, of course. One person may rely indirectly on another's data if he relies on the other person's testimony. The validity of someone else's data may thus be relevant by virtue of the condition that the lemmas be true. But if this condition is satisfied, empiricists assume that the *sufficiency* of a person's data is not affected by information someone else has. In making this assumption, empiricists overlook the social aspect of knowledge.

Suppose that Tom enters a room in which many people are talking excitedly although he cannot understand what they are saying. He sees a copy of the morning paper on a table. The headline and main story reveal that a famous civil-rights leader has been assassinated. On reading the story he comes to believe it; it is true; and the condition that the lemmas be true has been satisfied since a reporter who witnessed the assassination wrote the story that appears under his by-line. According to an empiricist analysis, Tom ought to know the assassination had occurred. It ought to be irrelevant what information other people have, since Tom has no reason to think they have information that would contradict the story in the paper.

But this is a mistake. For, suppose that the assassination has been denied, even by eyewitnesses, the point of the denial being to avoid a racial explosion. The assassinated leader is reported in good health; the bullets are said, falsely, to have missed him and hit someone else. The denials occurred too late to prevent the original and true story from appearing in the paper that Tom has seen; but everyone else in the room has heard about the denials. None of them know what to believe. They all have information that Tom lacks. Would we judge Tom to be the only one who knows that the assassination has actually happened? Could we say that he knows this because he does not yet have the information everyone else has? I do not think so. I believe we would ordinarily judge that Tom does not know.

This reveals the social aspect of knowledge. The

evidence that a person has is not always all the evidence relevant to whether he knows. Someone else's information may also be relevant.¹⁰ But how, exactly, ought the empiricist analysis be changed? Should we count information that any person at all has? Should we combine information possessed in part by several people, even if the information each has does not appear significant taken by itself? Must we take all the information one of these others has, or can we select bits and pieces that may give a misleading impression? And what is it that makes another person's information relevant?

The last question seems easiest to answer. Another person's information is relevant if the original person could not have properly reasoned as he did had he known about this information. If Tom had known about the denials as everyone else in the room knows, then Tom could not properly infer that the newspaper story is true. The other questions I have mentioned are not as easily answered, if we are to avoid the consequence that people rarely know anything. For example, if one could select bits and pieces of someone's information in a misleading way, he might be able to undermine almost any claim to knowledge. A similar result would follow if he could combine the information that several people hold separately,

since he might choose people such that their information combined to give a misleading result. On the other hand, it is not required that one combine the information everyone has, in order to see whether that prevents Tom's inference. That information would support Tom's inference, since it includes the fact that the explanations Tom originally inferred are correct.

The hardest problem is who may have the information that undermines Tom's reasoning. I doubt that we can allow his reasoning to be faulted by *any* one person's information. Otherwise, I would prevent many people from knowing things if I were to fake evidence about various things and show it to you. But I do not know how many people or what sort of people must be taken into account. Perhaps we must even consider people living at a different time, since we think our predecessors were sometimes right for the wrong reasons. It isn't just a matter of numbers. There can be evidence known only to a few that contradicts what the majority believe. This is certainly a subject worth pursuing; but I shall follow it no farther at this time.

In this paper I have tried to show two things. One is that there is something importantly right about the empiricist analysis. The other is that the analysis is not enough.¹¹

Princeton University

Received March 13, 1967

¹⁰ Why "social"? Can there be relevant evidence no one knows, has known, or will know about? I doubt it. In the example it is important that people have heard the denials. If they had been spoken into a dead microphone, I believe Tom would not be deprived of knowledge in the way he is by everyone's knowing about the denials.

¹¹ Apparently the social aspect of knowledge fails to provide a counter-example to the empiricist analysis of knowledge. Suppose we represent that aspect by the claim that the following condition is necessary for knowledge, where the condition is stated quite roughly and where we agree that there are serious problems in giving a precise formulation of the condition.

(1) No further evidence exists that would, if known, cast doubt on one's conclusion. Ernest Sosa mentions a similar condition in his article, "The Analysis of 'Knowledge That P'," *Analysis* 25.1 (1964), pp. 1-8 (see condition (oj₃)). Sosa also mentions another condition (sj₆) which I would express as follows:

(2) One must be justified in not believing that (1) is false.

To account for (2) we need only assume that the inference on which belief is based (if nondeductive) requires (1) as premiss or lemma. Furthermore the social aspect of knowledge then becomes a special case of the condition that the lemmas be true. Therefore, the social aspect of knowledge does not provide a counter-example to the empiricist analysis, indeed it is even to be explained in terms of that analysis along with (2).

IV. THE STATUS OF THE GENERALIZATION PRINCIPLE

SID B. THOMAS, Jr.

THE Generalization Principle states that what is right (or wrong) for one person in one set of circumstances is right (or wrong) for any similar person in similar circumstances.¹ I shall argue here (1) that this is a necessary principle of morality; (2) that it is also a substantive principle, in the sense that it can be used to establish moral judgments that could not be established without it; (3) that the *a priori* objections against its being both necessary and substantive are not cogent. Establishing these three points will occupy, respectively, the three sections of the paper.

I

The idea that a moral judgment about a particular act in a given set of circumstances is generalizable to similar acts in similar circumstances can be viewed as one of its features *qua* moral judgment.² This is the position defended by R. M. Hare.³ As he notes, it is a position about the "logic" of moral language; and he thinks that because it has this meta-status, nothing of substance can be derived from it (p. 31f). I accept his defense of the meta-thesis (as restated somewhat freely below), but shall add to it the point that a truth about the use of moral language can be transformed into a moral truth when the language is used rather than mentioned, and that this derived

moral truth (the G.P.) can then be applied to all sorts of cases.

I shall state Hare's case for the universalizability of moral assertions, as follows: Moral judgments about particular acts (e.g., "You ought not to smoke in this room") look as if they attribute some special kind of characteristic to the act ("moral required-ness"). Reflection on the situations in which these judgments are used, however, reveals two things: (a) there are no special moral characteristics to be found, either empirically, "in" the acts themselves, or conceptually, in a realm of Platonic Forms: moral judgments have a prescriptive, or action-guiding use, rather than a descriptive, fact-stating use, and (b) moral judgments are based on knowledge or belief about the non-moral features of the act, in the sense that the knowledge or belief that the act has these features provides the reason for making the judgment. (E.g., "There is an inflammable substance in this room which might be ignited by a spark from your pipe.") It follows from these points that the "moral characteristics" attributed to acts are supervenient upon non-moral ones: two acts cannot be known or believed to be identical in all non-moral respects and still be consistently thought to differ morally (p. 16ff). And from this follows what Hare calls the "logical thesis of universalizability": a person who uses moral words to make a judgment about a particular act

¹ The term "Generalization Principle" (sometimes abbreviated "G.P.")—as well as this formulation of the principle—is taken from Marcus G. Singer's *Generalization in Ethics* (New York, 1961, abbreviated "G.I.E." below). My agreement with Singer's view on the status of this principle, namely, that it is both necessary and substantive, will be evident to anyone who has read ch. II and III of his work. However, I have thought it better to develop the position independently, with only this general acknowledgment of my debt, both because it would have been tedious to try to link up the points I make with those he makes, and because I have my eye on some of his critics whom I would like to answer without having to work within the limits of the resources he provides. Cf. notes 7 and 11 below.

² Whether this "generalizability" is viewed as a distinctive feature of moral judgments, or as a feature which accrues to them by virtue of their being a type of "ought" judgment, is not at issue here. In either case it would be a necessary feature. For R. M. Hare's view on this point, cf. *Freedom and Reason*, sec. 3.3. (References in the paper are to the Galaxy Book edition of this work, Oxford University Press, 1965.)

³ In *Freedom and Reason*, Pt. I, ch. 2 and 3.

"commits himself thereby to a universal rule" (p. 30)⁴. (E.g., one who says "You ought not to smoke in this room" and gives the reason "An inflammable substance is present, etc.," commits himself to the universal rule "One ought not to smoke in any room in which inflammable substances are present."⁵ This is a logical thesis in the sense that it is a thesis about the way moral words and moral judgments function. Because it is merely a thesis about words, however, it by no means follows, Hare claims, that it is irrelevant to actual moral judgment-making. On the contrary; it has "great potency in moral arguments" (p. 31). According to it, one who wishes to prescribe what to do in a particular case cannot couch his prescription in moral terms without committing himself to having everyone act in the prescribed manner in all similar cases;⁶ and since this universal prescription would apply to himself as well as others, he cannot fail, if he is at all serious, to weigh his moral judgments carefully. There is no escaping the universal rule; a person is not free to "make an exception in his own case." If he states his prescription in moral language, he does not have the option of deciding whether it will apply universally or not. His only choice is whether to use the language of "ought" in the prescription.

As indicated above, I find this much of Hare's position satisfactory.⁶ There is, surely, a critical difference between the fact-stating and the action-guiding use of language—between "ought" and "is," if you like—notwithstanding the fact that certain terms and judgments may be used to serve both functions. And if this is so, then the very idea that there are special moral characteristics, either *in res* or *ante rem*, which are attributed to acts by

moral judgments, is essentially confused. No statement which merely attributed a characteristic to an act, of whatever sort, could by so doing serve a prescriptive function. This point simultaneously sweeps away the problem of the ontological status of moral "facts," and shows that the ought-is distinction can be secured quite independently of the narrow empirical bias which holds that only the sense-given properties of things are "real." On the other hand, it will hardly be denied, I think, that moral judgments are supervenient upon "is" judgments in Hare's sense. To deny this would amount to holding that it is possible actually to believe that an act ought or ought not to be done without knowing or believing anything whatsoever about it other than this—that it was the act of some human being. No one, I think, is likely to hold such a view. But from the fact that moral judgments are supervenient—tied to some features of the acts judged about, without being used for the purpose of stating that the act has those features—the thesis of universalizability follows. And again, from the fact that moral judgments are universalizable, it follows that a particular kind of *ad hominem* consideration is always relevant whenever someone makes one, namely, "how would you like it if what you hold to be morally required in this case actually were practiced by everyone in similar cases?"

The Generalization Principle also follows. What has been established so far, if the preceding argument is correct, is that one cannot say (e.g.) "A ought to do X in circumstances C" without committing himself, logically, to saying "Anyone like A ought to do X in circumstances C." But if this is true, then it must also be true, that if A ought

⁴ Even though Hare places universalizability and prescriptivity side by side as two distinct and independent aspects of his analysis of moral discourse (cf. p. 47, and ch. 6), the latter is actually needed to provide a theoretical basis for the former. If moral judgments did attribute some special kind of (non-natural) characteristic to acts, the connection between the morality of the act and its "circumstances" would be merely inductive and contingent. It would be (logically) possible for two acts to be identical in all non-moral respects and yet differ morally, which is precisely what the universalizability thesis denies. The prescriptive character of moral judgments, however, prevents this possibility, and is hence needed to establish universalizability.

⁵ Strictly speaking, a *ceteris paribus* clause is needed here; but I shall not be concerned with this complication.

⁶ For a defense of the universalizability thesis which does not depend on the "Prescriptive" interpretation of ought-judgments, cf. C. D. Caton, "In What Sense and Why 'Ought'-Judgments are Universalizable" (*Philosophical Quarterly*, vol. 13 [1963], pp. 48-55). The argument given by Caton is this: (1) first-level "ought" judgments are logically the kind that must be supportable by reasons (p. 50); (2) reasons are always universalizable ("whenever a person offers a reason for something he offers a reason adaptable to any case which is not dissimilar in some relevant respect" [p. 51]); (3) therefore, etc. This argument is sound, but since he does not avail himself of the idea that moral judgments are essentially prescriptive (rather than fact-stating), he has no adequate explanation of why (1) is true. He argues simply that if a person doesn't understand that it is legitimate to ask "why?" when a first-level ought judgment is made, "we would begin to doubt that he was using 'ought' in a way with which we are familiar." This makes it seem as if it might be just a contingent fact about our language that such judgments are supportable by reasons, whereas, if their action-guiding character is taken into account, one sees that such support is required by anyone who does not wish to base his action on (his or someone else's) whim.

to do X in circumstances C ; anyone like A ought to do something like X in those circumstances. And this is simply another version of the Generalization Principle. The G.P. is merely the thesis of universalizability restated in such a way that moral terms are used rather than mentioned. The same argument that establishes the implicit universal character of particular moral judgments also establishes the G.P. If it is self-contradictory to say " X and T agree in their non-moral characteristics, but X is right and T is wrong," then X cannot be right and T wrong when they so agree. Moreover, since the denial of the Generalization Principle would be self-contradictory, it is not only true but necessarily true.⁷

Hare himself would not approve of the turn the argument is taking. He appears to think that because universalizability is a logical feature of moral judgments it cannot be used to substantiate particular moral judgments. Thus he says: "The logical thesis has, as we shall see, great potency in moral arguments; *but for that very reason* [my emphasis] it is most important to make clear that it is no more than a logical thesis—for otherwise the objection will be made that a moral principle has been smuggled in disguised as a logical doctrine" (p. 31).⁸ What if the objection were made? It would carry force only if the person making it assumed the impossibility of deriving anything substantive from a purely "neutral" thesis about the logic of moral language. The fact that Hare evidently thinks such an objection would carry force, and tries to meet it by showing that his thesis carries no substantive commitments, shows that he shares the assumption. It is precisely this assumption, however, which is false, as I shall show in the next section. The Generalization Principle, which follows directly from the universalizability thesis, does have important substantive uses.

II

In order to understand clearly what these uses are, it is best to emphasize at the outset that the

G.P., by itself, cannot yield any conclusion about what particular acts are right (or wrong). It merely states that an act cannot be right (or wrong) for a given person in given circumstances without being right (or wrong) for similar persons in similar circumstances. Its use in moral reasoning, therefore, presupposes other principles. But when used with other principles, it does have a "substantive" role to play in requiring that they be applied consistently to similar cases, even when the factual data relating to the persons and circumstances in question may be in some respects ambiguous. Its special use as a distinct moral principle comes to the fore mainly when there is some kind of determinancy about the cases to which the other principles apply which makes it difficult or impossible to decide those cases precisely on their own merits. When this occurs, the G.P. says that all must be decided alike. It says that the decision in one case cannot be correct unless that decision were correct in every similar case.

In order to make this use of the G.P. clear, consider a case of punishment. Suppose that Smith and Jones both violate the same law or regulation for essentially the same reasons; that they both confessed; and that neither can appeal to any extenuating circumstances to lessen his penalty that is not also available to the other. The question of what their punishment ought to be is not one that can be decided by referring to the G.P. It would have to be determined by some other principle. But the problem of "fitting the punishment to the crime" is notoriously difficult, whatever principle is applied. There is a range of indeterminacy here. Would X number of days' imprisonment be justified? Perhaps. $X+1$? Perhaps that, too; it may well be impossible, in principle, to decide between them in such a way that one could say with assurance "Exactly X days of imprisonment is a just punishment in this case: no more and no less." So if Smith's case and Jones's case were considered separately on their own merits, *different but equally justified* punishments might well be assigned. But this is precisely what the G.P. prohibits. It says that the assignment of X days of imprisonment in

⁷ For another, but significantly different, way of stating this point, cf. Singer, G.I.E., p. 47. The main difference between my defense of the G.P. and Singer's is that I base its necessity on the prescriptive and supervenient character of moral judgments, whereas he, like Caton, argues directly from the fact that moral judgments are necessarily supportable by reasons, which are essentially generalizable. (Cf. Ch. 3, sect. 1.) Thus, he also has no explanation of *why* reasons are required (cf. note 6 above); in fact, he has even less basis for his claim, since he refuses to support the G.P. by appealing to the idea that it enshrines a fact about the use of moral language (cf. p. 46ff.), while still claiming that it is "a principle so fundamental as to be a pervasive feature of moral language." (P. 49.) On what basis, then, does the G.P. rest for Singer? Is he offering us a new kind of intuitionism?

⁸ Cf. also sect. 10.1.

Smith's case cannot be correct unless the assignment of that number of days in Jones's case is correct. It requires the judge, or parent, or whom-ever, to generalize the decision from one case to the other, even though independent consideration of that other case might have resulted in a somewhat different, though—on the basis of its merits alone—equally justified decision.

It is not relevant to point out that if the cases of Smith and Jones are considered *ad seriatum* rather than simultaneously, the "judge" may make a mistake in the first one, and assign a punishment that was too harsh or too easy which he wishes to correct in the second case. The G.P. would not prohibit him from making the correction, because it wouldn't apply. If he made a mistake in the first case, the decision is not required to be extended to the second; this is only required when no mistake was made. The point of the G.P. is to impose a condition for deciding *correctly*. It states that both cannot be decided correctly if they are decided differently.

It may be noted further that if the cases should be decided differently, the person receiving the heavier punishment would have a legitimate complaint against the judge, based squarely on the G.P. This complaint could not be turned aside by saying that the different punishments were both equally justified on the basis of a separate consideration of the merits of each case.

The Generalization Principle can be applied in the same way to justify the right to equal salaries for equal work. As in the case of punishments, it may be difficult, if not impossible, to determine, to the very dollar, what a fair wage would be for a given piece of work. Whatever principle is used will surely leave an area of indeterminacy here. The G.P. cuts across this area of indeterminacy and flatly requires that a salary cannot be justified in one case without being justified in all similar cases. It explicitly prohibits making a separate decision on the merits of each case alone, and attempting to justify the discrepancy by appealing to the inexactness of the notion of a "just wage."

A somewhat different case of the substantive use of the G.P., which does not involve settling a range of indeterminacy in quite the same way, concerns the granting of privileges. Everyone is surely familiar with the child's complaint: "If you let him,

you've got to let me." It is common, relevant, and one that ought to be taken into consideration by anyone who is deciding whether to "let him" or not. It carries force even when the "letting" is an essentially neutral thing, from the moral point of view—that is, when the asker has no special moral claim to press. Consider, for example, the situation of a public official who has the keys to the local high school gymnasium, and the authority to let non-students use it at off times. If he takes it on himself to let some of his friends do this, he must be prepared to deal with the case of others who hear about it and present themselves as candidates for the privilege. They will have a moral argument, again backed squarely by the Generalization Principle: "If you let them, you ought to let us; and you *did* let them." He can only reply (a) by admitting straightforwardly that he is going to exclude them *arbitrarily*—which leaves their moral claim quite intact;⁹ or (b) by admitting that he ought not to have let his friends in, and announcing that no one will use it hereafter—which again grants their point; or (c) by letting others use it too. There is no way around the force of their argument. And it is the G.P. alone which gives it this force. The second group has no *independent* claim to press. By hypothesis, the merits of their case, considered by itself, are nil. But by letting the others use the gym, the official is no longer to decide the second case on its own merits. Its "merits" now, by the G.P., include the way the other case was handled. He could not have been acting correctly, from the moral point of view, in letting his friends use it, unless it would also be correct to let the others use it; that is what the G.P. says. Hence he must either announce that he is not going to take moral considerations into account at all in his action or carry through in one of the two ways just indicated. The G.P. here gives rise to a moral requirement to either terminate the privilege the first group enjoys or extend it to the second.

It may perhaps be thought that, in dealing with these cases, I have applied a principle other than the Generalization Principle, and hence have not shown that the latter has a substantive use. Thus it may be objected that while the G.P. states that the punishment allotted Smith (to go back to that example) cannot be justified unless the justification would hold for allotting Jones the same punishment it does not follow that they ought to be punished

⁹ Note that I am assuming that "being a friend of X's" is not a relevant difference in this case. In general, I am skirting the issue of how relevant differences are determined, for it concerns the application of the G.P. rather than its status as a principle. On what does and does not count as a relevant difference, cf. Singer, ch. IV, sects. 3–5, and Nakhnikian's criticism in "Generalization in Ethics," *Review of Metaphysics*, vol. 17 (1964), pp. 449ff.

alike. For where there is a range of indeterminacy with regard to justifiable punishments, it is one thing to show that the same penalty for both could be *justified*, but another thing to show it is *required*. According to this objection, the principle needed to yield the requirement that they be punished the same is something like "treat like cases alike"; and, it might be urged, this is not equivalent to the G.P.—"what is right for one is right for all"—even though they are quite similar and would coincide whenever the justification for an act would yield the actual requirement that it be done.

My reply to this is that the distinction between what is justified in two relevantly similar cases and what is required in them cannot be drawn here, precisely because the G.P. prohibits it. Where there is indeterminacy, one can say with regard to the individual case that there is a distinction between what is justified and what is required; that would in fact determine "indeterminacy." But when the second, similar, case is considered, the treatment that was justified in the one is the treatment that is *required* in the other. For if relevantly similar cases are treated differently, then by the Generalization Principle, both cannot be treated correctly. What is morally required, however, is that neither be decided incorrectly, or unjustifiably. Hence, the G.P. requires that they be treated the same. As I said above, the G.P. determines a condition for deciding similar cases correctly.

III

I claim then, (1) that the Generalization Principle is a necessary moral truth, being simply the object-language correlate of the universalizability thesis; and (2) that it has a "substantive" use in moral reasoning over and above the *ad hominem* use that Hare allows.

From a strictly logical point of view, however, it seems impossible for both these claims to be correct. In saying that the Generalization Principle has a substantive use, I mean it can be used, in the kinds of situations illustrated above, to establish synthetic moral judgments that could not be established without it. But if this claim were true, it would seem as if it could not be a necessary moral principle. It seems logically impossible for the truth of a synthetic statement, of any sort, to depend essentially on the truth of a necessary statement.

This *a priori* difficulty rests on the assumption

that moral language is decriptive, or has an essential fact-stating use. If this assumption is dropped, the objection no longer carries any weight. For to drop this assumption is to drop the idea that the correctness of a moral judgment depends on the existence of some independently constituted "fact"; and the distinction between necessary and contingent truths depends upon just this idea.

Let me try to explain this more clearly. In the case of language having a fact-stating use, we tacitly suppose that if a certain object has the property *P* and also has the property *Q*, where *P* and *Q* are distinct, it cannot be validly inferred that another object having *P* must necessarily have *Q*. Indeed, this is ordinarily taken as a criterion for the "distinctness" of *P* and *Q*.¹⁰ The operative idea here is "whatever is distinguishable is separable." Again, on the basis of the same assumption, we suppose that if the inference is valid—if "*y* is *Q*" can be validly inferred from "*y* is *P*" (where "*y*" is anything whatever), given only the premiss that "*a* is *P* and *a* is *Q*"—then *P* and *Q* cannot be distinct. If they were distinct, we assume, they would be separable, and there would be no basis for the inference. It would not be self-contradictory to suppose that some *y* is *P* and not *Q*. The idea that there must be a *basis* for the inference from *P* to *Q*, and that this basis can only be rendered intelligible through appeal to the law of non-contradiction and the notions of identity and difference is, I think, perfectly correct with regard to fact-stating language, and might even be said to define its domain. But where the assertion that an object "is *Q*" does not play a fact-stating role, the situation is radically altered. What the assertion "attributes" to the object in this case is not an objective property at all, and hence the question of whether *P* and *Q* are "distinct and separable," or "distinct but not separable," does not arise. This is precisely the case where *P* is a non-moral and *Q* a moral predicate. *Q* here is the sort of term that can only be *known* to be attributable to an act by knowing it has some non-moral property *P*; but the dialectics of the "distinguishable is separable" principle do not apply because it is incorrect to say that *P* and *Q* are either distinct or not distinct. The inference from "*a* is *P* and *Q*" to "if *y* is *P*, then *y* is *Q* (whatever *y* may be)" requires no basis in the law of non-contradiction. Because moral language is supervenient upon non-moral language, no independent criterion for the truth of similar but non-identical

¹⁰ The difficulty in applying this criterion need not concern us here.

moral "facts" is possible.¹¹ One might say that with respect to moral language, there is no distinction between the "fact" expressed and correct use of the language which expresses it.

There is another *a priori* objection to the thesis I am maintaining which deserves careful consideration. D. H. Monro has argued that anyone who holds, as I do, that "universalizability is a purely formal feature of moral terms," and then proceeds, as I do, to try to derive from it a substantive moral principle, must be confusing the *Principle of Impartiality* with the *Principle of Consistency*.¹² The latter is stated in this way:

PC: If *R* is a reason for *X* to do *Q*, it is (i) a reason for *X* to do something like *Q* in similar circumstances; (ii) a reason for someone else, *Y*, who is like *X* in the relevant respects, to do something like *Q* in similar circumstances.

This may be a logical principle, he admits, in the sense that it is "analytic by virtue of the meaning of the word 'reason'" (p. 165); but (a) necessary conformity to it is not a distinguishing feature of moral judgments, since any practical judgment which was advanced on the basis of reasons must necessarily conform to it, whether the reasons are moral or non-moral; and (b) it is "utterly different" (p. 166) from the Principle of Impartiality:

PI: One ought not to make exceptions in one's own favor.

The last point is crucial; the first may be readily granted, since it does not at all tend to show that moral judgments are not essentially universalizable. (cf. note 2 above). In the context of Monro's discussion, it is perfectly clear that in saying the PC is "utterly different" from the PI, he means to say the latter cannot be deduced from it. Thus he insists,

¹¹ This effectively replies to one of Gewirth's main criticisms of Singer (Alan Gewirth, "The Generalization Principle," *The Philosophical Review*, vol. 73 [1964], pp. 229-242). He urges that either there is no way of determining the rightness (or wrongness) of a particular act, independent of the G.P.—in which case the latter is a tautology and cannot make any substantive moral difference; or there is a way of determining the moral nature of particular acts independently—in which case it cannot express an aspect of the "logic" of moral discourse to which every judgment must conform. Hence, the G.P. either makes no substantive moral difference or it cannot express an aspect of the logic of moral discourse (pp. 231-235). There is an ambiguity here in the notion of an "independent" principle. There are (in fact there must be—cf. p. 5 above) principles independent of the G.P., in the sense that they can be applied to particular acts without explicitly raising the question of whether the moral judgment being pronounced would be correct for every similar case. But these principles are *not* independent of the G.P. in another sense, for the latter belongs to the logic of the kind of judgment being considered. Gewirth's dilemma neglects the special nature of moral judgments which makes it possible both for the G.P. to be a tautology (because they are supervenient upon non-moral judgments) and for there to be criteria *other than it* for applying moral predicates. His argument is appropriate to fact-stating language, where there is a distinction between the truth of the assertion and the correct use of the language in which it is made. No such distinction exists in the case of morals. Having failed to develop a theoretical basis for the G.P. (cf. n. 7 above) Singer is not in a position to make this reply to Gewirth.

¹² "Impartiality and Consistency," *Philosophy*, vol. 36 (1961), pp. 161-176. References in the paper are to this article.

wrongly as I shall try to show, that the Principle of Impartiality is synthetic; it can be denied without contradiction. Anyone who asserts it, he says, "is taking a moral stand" and "the issue between someone who asserted it and someone who denied it would be a moral issue." From this alone, if it were true, it would appear to follow that the PI cannot be deduced from the PC; for the latter is analytic.

But it is not true. The Principle of Impartiality just is the Principle of Consistency "instantiated" for moral reasons, and hence must be a necessary moral principle. The result of inserting the word "moral" into the PC in the appropriate places is:

MPC: If *R* is a moral reason for *X* to do *Q*, it is (i) a moral reason for *X* to do something like *Q* in similar circumstances; (ii) a moral reason for someone else, *Y*, who is like *X* in the relevant respects, to do something like *Q* in similar circumstances.

And this is simply another version of the Principle of Impartiality. "To make an exception in one's own favor" means to claim oneself exempt from a moral requirement admitted to be binding on others, without referring to any relevant differences to justify the exemption. The MPC, then, is equivalent to the PI; and since the former is logically derivative from the PC, an admitted necessary truth, the latter must be necessary also.

This merely sets the stage for what I think is Monro's most important argument, which I shall re-state as follows: A distinction must be made between two uses of the term "moral": it has "a descriptive use, in which it is contrasted with 'unmoral' or 'non-moral'; and an evaluative use, in which it is contrasted with 'immoral'. In the first of these it makes sense to talk of the moral policy (or 'morality') of Satan; though here again, when

we think of the other use, we are tempted to protest that this isn't a morality at all, but the denial of all moral principles. And here again, 'there are no bad moral principles' is, in the evaluative sense, an analytic statement; in the descriptive sense, it is synthetic (and false)" (p. 161). Now if it *could* be shown that the PI is a necessary moral principle in the evaluative sense, it would still be the case that it had a significant denial which was a moral principle in the descriptive sense. As he says: "It is hard to see how the omission of 'not' [from the PI] could transform a moral principle into a non-moral (as distinct from an immoral) one" (p. 163). Thus the only kind of "necessity" the Principle of Impartiality could conceivably have would be a relative one. It may be a necessary principle for anyone who wishes to make moral judgments in conformity with the Principle of Consistency, but it is not necessary for someone to choose not to do that; and hence the PI is not an absolutely necessary principle of any morality. It is *possible* to give moral approval to the principle "let a person make exceptions in his own favor." Perhaps we might say a person who gave such approval has a mistaken morality, but then we should be evaluating his moral principles on the basis of our own; we should not be denying that he had any moral principles at all. The PI may represent a very fundamental moral commitment for us (it is, he says "fairly fundamental, at least in our culture"—p. 176), but the fact that it does represent a *moral* commitment (evaluative sense) shows that it cannot be used to rule out its opposite as a possible moral principle. A *substantive* moral principle cannot be "put forward as enabling us to distinguish moral terms (and moral judgments) from others" (p. 162; cf. also p. 166).

This position, I think, is demonstrably wrong. The precise point at which it goes wrong, however, is not evident, and requires some spadework to uncover. I shall first give my reason for saying it is wrong, then turn to the digging.

As Monro himself points out in another place (p. 166), the Principle of Consistency defines the difference between deliberate and arbitrary behavior. One who acted without reasons would be one who acted arbitrarily, or at random. Behavior governed by moral considerations, on the other hand, cannot be purely arbitrary or random; if *any* way of acting was as morally acceptable as any other, that would be tantamount to saying that no moral distinctions exist at all. It follows that morally governed behavior is based on reasons—it

falls under the class of deliberate behavior. And by the Principle of Consistency, it follows that "a reason for one is a reason for all"; and from this, that "what is right (or wrong) for one is right (or wrong) for all." And this means that the Principle of Impartiality is a necessary principle of any morality at all, even the descriptive sense. Indeed, it should be obvious that if the denial of the Principle of Impartiality were a possible moral guide, there would be no such thing as moral guidance. If everyone could claim the right to make exceptions in his own favor, this would render the effect of any proposed moral requirement null and void. This shows that the attempt to put forward the denial of the PI is really self defeating. Its adoption would make the notion of an "exception to a moral requirement" inapplicable, since there would be no case which was not an exception.

Now for the digging. The source of Monro's error lies, I believe, in the use of an inappropriate model for interpreting the relation between the evaluative and descriptive senses of "moral." It would seem that the descriptive sense is generic: the distinction between "moral" and "non-moral" judgments is logically prior to the distinction between "moral" and "immoral" judgments, as the distinction between "animal" and "non-animal" is logically prior to that between "two-legged animal" and "non-two-legged animal." From this model, it would follow that the denial of any moral judgment, in the descriptive sense, is another moral judgment—in the descriptive sense: just as non-two-legged animals are as much animals as two-legged ones. And from this it follows that no moral judgments in the *normative* sense could possibly furnish a criterion for distinguishing moral from non-moral judgments. That would be tantamount to supposing, e.g., that an animal had to be two-legged in order to be an animal at all, after admitting that it was possible for there to be non-two-legged animals.

But this genus-species way of viewing the matter is wrong. The distinction between correct and incorrect moral judgments is logically prior to the distinction between moral and non-moral judgments, although in a sense of course it is true to say that correct moral judgments are a species of moral judgments in general. The point is that the possibility of identifying a judgment as moral in the descriptive sense depends upon the possibility of identifying *correct* moral judgments. Nor is the language of morals unique in this respect. The

distinction in logic between deductive and non-deductive arguments is similarly related to the distinction between correct and incorrect deductive arguments. The possibility of identifying an argument as "deductive" is logically dependent upon the possibility of identifying *correct* deductive arguments.

Let me elaborate this latter example more carefully. Since it is exactly parallel to the case of moral judgments it will be easy to transfer the results from one domain to the other.

Consider the case of someone who might argue with regard to the deductive-non-deductive distinction as Monro does for the moral-non-moral distinction. "Surely," he would say, "there is a *descriptive* sense of 'deductive argument' in virtue of which an argument can be identified as 'deductive' without prejudging its validity. If this were not so it would follow that every argument correctly termed 'deductive' is valid, and there could be no such thing as a 'formal fallacy'. This is absurd. Hence there must be some criterion for distinguishing deductive from non-deductive arguments, and this must be logically independent of the criterion used for determining whether a given deductive argument is valid." This reasoning is exactly parallel to Monro's, but here the fallacy is easily detected. Fallacious deductive arguments are of course possible; but this does not imply that there is a way of identifying deductive arguments in general—fallacious or non-fallacious—which is logically independent of the criterion for identifying non-fallacious deductive arguments. A fallacious deductive argument is such only because it makes a claim to be something it is not—a valid deductive argument. It, or the person advancing it, claims that the relationship between premisses and conclusion is that of logical entailment. If no such claim were made, the argument could not fail in the way it does fail. Hence the idea of *claiming to be correct* is essential to the notion of *any* deductive argument, fallacious or non-fallacious. It is impossible to determine whether any argument is a deductive one without first determining what claim is being made for the relationship between premisses and conclusion.

But on the other hand, it is possible to determine whether a given argument is a valid deductive argument without considering such a claim. To determine whether it is valid or not is to determine whether the relationship between premisses and conclusion is that of logical entailment; and this can be done by applying the criteria of validity.

Whether the relationship does in fact hold does not in any way depend upon whether the claim that it holds has been made.

Here, then, we encounter an apparent anomaly: whether an argument is a *valid* deductive argument can be determined without even raising the question of whether it is a deductive argument at all. This violates the apparent truism that membership in a species cannot be determined without first, or simultaneously, determining membership in the genus. Or, I would rather say, it shows that the genus-species model for understanding the relationship of "deductive argument" to "valid deductive argument" is not applicable. This shows the reason why the distinction valid-invalid is more fundamental from the standpoint of *identification*: if it were impossible to make it, it would be impossible to make the other. If nothing counted as a correct deduction, then nothing could count as a claim to be a correct deduction, and nothing could count as having failed to make one. But, as we just pointed out, the claim to be a valid deductive argument is essential to the application of the notion of "deductive argument" in the descriptive sense. Hence the possibility of applying the deductive non-deductive description depends, logically, on the possibility of applying the correct-incorrect distinction.

All this applies straightforwardly to the claims I have made for the Generalization Principle. To make a moral judgment about behavior is to make the claim that the judgment is based on reasons—that it is based on some facts about the case; so that if the facts were different, the judgment might be different too. This claim is as essential to the notion of a moral judgment as the claim to validity is to the notion of a deductive argument. If the claim were explicitly retracted (if it were said that there are no reasons for the judgment at all) then nothing could count as making a correct moral judgment, since by hypothesis, there would be no reason why it rather than its contradictory would not be correct. Hence nothing could count as failing to be correct. But if no one could fail to be correct in making a moral judgment, there would be no point in making them at all; any judgment made would be the "right" one. If, then, there is any point in making moral judgments at all, they must be assumed to be based on reasons—whether the person can actually give the reasons or not. To reject the Generalization Principle, however, is to reject the possibility of being moral judgments on reasons. Hence, if there is any point in making moral judgments, the G.P.

cannot be rejected. One who tried to reject it, and put forward its denial as a "moral" principle, would be defeating himself. He would be claiming that there was some point in making moral distinctions, while denying a necessary condition for there being any point in making them.

If, however, the claims to be based on reasons is essential to the notion of a moral judgment, in the descriptive sense, and if this claim implies recognition of the Generalization Principle, then the possibility of identifying a judgment as *moral* (in the descriptive sense) presupposes the correctness of that principle. Thus Monro's argument fails. One does not in any sense have the choice of accepting the denial of the Generalization Principle as a feature of his "morality"; he only has the choice of

being moral or not. The principle is not just fairly fundamental "in our culture"; it is a necessary principle of any morality in any possible culture.

I should perhaps add that the above argument in no way commits me to the position that there are principles other than the G.P. which one might use in distinguishing correct from incorrect moral assertions which are derivable from the logical features of moral discourse. If there are such principles, well and good. The criterion would be whether it was possible to reject them without entirely rejecting the possibility of making moral distinctions (in the descriptive sense), and that would have to be argued separately for each case. For the case of the Generalization Principle, the argument is, I think, conclusive.

State University of New York at Binghamton

Received May 2, 1967

V. WHAT IS AN EPISTEMOLOGICAL PROBLEM?

JOHN L. POLLOCK

I. THE FUNDAMENTAL PROBLEM OF EPISTEMOLOGY

LET us begin by looking at a classical example of an epistemological problem:

There is a book sitting on my desk in front of me. But now, suppose I ask myself how I know that there is, or more generally, how I know that there is anything there at all (regardless of whether it is a book). A sensible answer to this question would be, "Because I see it." We know that there are material objects around us because we see them, feel them, hear them, etc. And the statement that we see something, or feel it, or hear it, logically entails that it is there to be seen, or felt, or heard. But now, I *say* that I see something (a book) there before me on my desk, but how do I *know* that I do? Mightn't I be hallucinating, or seeing an after-image, or witnessing some sort of cleverly constructed optical illusion? My experience might be exactly the same as when I really am seeing a book, and yet there might not be any book there, because I am hallucinating. Generalizing this, how do we know we *ever* perceive the things we think we do? Might we always be hallucinating?

As it has sometimes been denied that it is even meaningful to suppose we might *always* be hallucinating, let us make this possibility more concrete as follows: Suppose that a group of psychologists, bio-physicists, and neurologists have constructed an adequate explanation of the neurophysiology of perception, and to test their explanation they take a subject from birth and wire him into a computer which directly stimulates his brain in such a way as to give a coherent, but completely false, sequence of sensations. In the subject's own mind he would seem to live out a completely normal life, growing up, making friends, going to school, getting a job, marrying and raising a family, etc. And yet all those years he was really sealed into an experimental apparatus in which he was fed intravenously and never had any contact with the outside world. It is no doubt true that in the present state of neurophysiology

this could not be done, but it is certainly a meaningful hypothesis and a logical possibility.

Now, how do I know that I am not in the position of the subject of the above experiment? Perhaps a group of scientists have me hooked into such a computer, and all of the experiences that I think I have had since birth are really figments of the computer. This of course is an absurd possibility, but the interesting question is, how do I know that it is absurd? What sorts of grounds can I adduce for its falsity? It seems that any reason I can have for thinking that this possibility is false must either be a reason for thinking that it is logically impossible, or else must be an empirical reason, arrived at inductively, for thinking that it is false as a matter of fact. But it is hard to see how this sceptical hypothesis can be logically impossible (it seems to make perfectly good sense—we know what it would be like for someone to be so wired into a computer), and it seems that in order to have inductive evidence for an empirical reason I would already have to be able to rely upon some of my perceptions—which I cannot do without simply begging the question against the sceptical hypothesis. How then can I know that the sceptical hypothesis is false?

This is representative of the type of problems with which epistemologists have traditionally concerned themselves. The problems of epistemology are problems of how we can possibly know certain kinds of things that we claim to know or customarily think we know. In general, given the statement that *P*, we can ask, "How do you know that *P*?" This is the general form of an epistemological problem. The question "How do you know that *P*?" is a challenge—a demand for justification. The task of the epistemologist is to explain what justifies us in believing the things we do.

Let us call the conditions under which one would be justified in believing that *P* the *justification-conditions* of the statement that *P*. Then we can distinguish between two problems. The first is to *state* the justification-conditions for different kinds of statements, and the second is to *prove* that they are the justification-conditions. Of course,

these two problems do not seem to be unrelated, but it must be borne in mind that there are these two distinct problems. The latter problem has generally interested epistemologists more than the former problem. Epistemologists have usually been content to give only a very rough description of the justification-conditions of statements, and then have gone ahead to try to prove that those are the justification-conditions. For example, the classical "Problem of Induction" is one of justifying induction as a way of learning the truth of universal generalizations. Although they've never been very clear on just what those grounds are, few philosophers doubt that we do base knowledge claims on inductive grounds. But what is demanded is a proof that we are justified in doing so. And similarly, the "Problem of Our Knowledge of the Material World" is the problem of explaining how we can justify basing knowledge claims about material objects on sense perception. It is not doubted that we do in fact base them *somehow* on sense perception, but what is wanted is a proof that we are justified in doing so.

II. REDUCTIVE ANALYSES

The fundamental problem of epistemology is to explain what it is that justifies the kinds of knowledge claims that we customarily make. This problem has traditionally been construed as requiring a justification for our basing knowledge claims on the grounds on which we do in fact base them (a *proof* that what we suppose to be the justification-conditions really are the justification-conditions).

On the face of it, there seems to be a very good reason why we should, in principle, be able to give a proof of the sort desired. If we cannot establish any connection between one state of affairs and another, then we cannot be justified in making claims about the latter state of affairs on the basis of the former. Thus if we cannot justify our customary grounds for knowledge claims, then we cannot take them as justifying our claims to knowledge, and so they are not really good grounds at all. Therefore, unless we can, in principle, give a proof of the sort desired, we are led to scepticism.¹

So let us see how we might set about justifying our basing knowledge claims on some particular

source (such as perception). It seems that there can only be two ways in which this might be done. We could either justify it inductively, showing that it does in fact tend to lead to true knowledge claims, or else we could justify it logically, showing that there is some sort of logical connection between the source and the knowledge of which it is a source. But an inductive justification is impossible. We could only inductively justify a source of knowledge if we had independent access to both the source and the knowledge of which it is a source, and then could compare them and see that there is a correlation. But we do not have independent access to the knowledge that these sources are supposed to provide. They constitute *the* sources of this knowledge. For example, we do not have access to the material world except *through* perception, and so there is no way to compare the material world with perception to see that perception is a reliable guide unless we beg the question and assume in the beginning that it is.

It seems then that the only way to justify a source of knowledge is by establishing some sort of logical connection between the source and the knowledge it is supposed to give us. A logical connection must arise from the meanings of the concepts or statements involved in the knowledge claims. And (and *here* is the step which I shall deny) it has traditionally been supposed that the only way to analyze the meaning of a statement or concept is by giving its truth-conditions—by saying what conditions must be satisfied in order for the statement to be true or for the concept to be correctly ascribable to an object.² Furthermore, not just any statement of truth-conditions will suffice. Starting from the truth-conditions of a statement, we could never establish a logical connection between that statement and the source of knowledge which is supposed to yield the statement unless those truth-conditions were stated in terms of the same concepts as are used in describing the source. Thus for example, we could never establish a logical connection between perception and statements about the material world unless we could state the truth-conditions of the latter in terms of perception.

An analysis of the truth-conditions of a statement in terms of the source of our knowledge of that kind of statement is what philosophers have called a *reductive analysis*. Since Descartes, who was

¹ I shall later contest at least part of this claim.

² We have inherited this assumption most recently from logical atomism and logical positivism, but it is in fact not new. It is really the same sort of thing that Locke, Berkeley, and Hume—and for that matter Socrates—were doing.

largely responsible for getting modern philosophy started on the problem of justifying sources of knowledge, epistemologists have been concerned almost exclusively with giving reductive analyses of statements, and we now see why. Given assumptions that philosophers have traditionally accepted, it follows that the only way to prove that the justification-conditions of a statement are the justification-conditions is by giving a reductive analysis of the statement in which the truth-conditions are stated in terms of the same concepts as the justification-conditions.

Philosophers have commonly supposed that they know more or less how we are justified in making the kinds of knowledge claims we do. Although they haven't generally been able to state the justification-conditions precisely, they felt that they could at least pick out the general sources of our knowledge in different areas. For example, our knowledge of the material world comes from sense perception. And our knowledge of other minds comes from people's behavior. And our knowledge of right and wrong (if we can properly speak of "knowledge" here) comes from non-moral states of affairs in the world. Thus if we are to justify these sources of knowledge, it seems we must seek reductive analyses of these statements in terms of these sources. Working within this traditional framework, phenomenalism becomes the only possible theory of our knowledge of the material world, behaviorism becomes the only possible theory of our knowledge of other minds,³ and naturalistic ethics becomes the only possible theory of our knowledge of ethical truth.

It seems then that in order to justify sources of knowledge we are driven inexorably to reductive analyses. This seems to be the only way to derive the justification-conditions of statements from the meanings of those statements. And it seems that if the justification-conditions are not derivable from the meanings of the statements—if there really is no logical connection between them—then they cannot be the justification-conditions, because if we cannot justify our sources of knowledge, then they do not justify our claims to knowledge, and so are not really sources of knowledge at all. It seems that we must either have reductive analyses or scepticism. They are the only two possibilities.

Let us ask then whether reductive analyses are

always possible. Unfortunately, the answer seems to be, "No." For example, consider our knowledge of the material world. A reductive analysis there would take the form of phenomenalism. I think the strongest argument against phenomenalism is that of "perceptual relativity." According to phenomenalism a material object statement is to be analyzed as a (perhaps infinite) conjunction of statements about experience. Then the material object statement must entail each of those statements about experience. But a material object statement cannot entail any statement about experience, because given any material-object statement and any statement about experience, we can imagine circumstances in which the material-object statement is true, and yet due to malfunction of certain organs, or hallucination, the statement about experience is false. For example, we can recall the example given in the previous section of the man who is wired into a computer which directly stimulates his brain. This is done in such a way that his experience is entirely independent of what is actually going on around him, so that his experience is always delusive.

Thus it seems that the search for reductive analyses which will solve the problem of our knowledge of the material world is stymied. Nor is this the only place in which the attempt to find reductive analyses seems to have failed. Consider our knowledge of other minds. Here the reductive analysts are the behaviorists who try to reduce statements about persons to statements about material objects. But behaviorism doesn't seem to work, and so reductive analyses don't seem to offer a solution to the problem. Consider the philosophy of logic. Quine and his followers have argued, in effect, that no reductive analyses can be given of such logical concepts as analyticity, implication, and logical necessity, and on this basis they have actually been led to embrace a kind of logical scepticism wherein they deny the very existence of these concepts. We have a corresponding problem in ethics. Reductive analyses in ethics take the form of naturalistic ethical theories in which the meaning of an ethical statement is analyzed in terms of statements about the material world. But the apparent existence of the Naturalistic Fallacy vitiates such analyses. In all of these areas the search for reductive analyses seems to have been fruitless.

³ This is oversimplifying a bit the case of behaviorism, because the possibility of an inductive justification of our source of knowledge has been maintained by the supporters of the Argument from Analogy. But if we can agree with those who maintain that the Argument from Analogy will not work, and thus that an inductive justification is impossible, then it seems to follow that behaviorism is the only possible theory.

In the case of some problems, such as the Problem of Induction, the hope that some reductive analysis might be found was just patently absurd. In the case of the Problem of Induction, a reductive analysis would involve stating the meaning of unrestricted generalizations in terms of finite combinations of singular statements, which no one *ever* thought could be done. This is just what has always made the task of justifying induction so puzzling. In order to show that induction was justified some philosophers have instead sought an additional premiss concerning the uniformity of nature. But even if such a principle could be found, in order to do the job that premiss would have to be necessarily true (it couldn't be justified inductively without circularity). And on these philosophers' own premisses, if it were necessarily true it would have to be derived from the truth-conditions of general statements. And that again would involve a reductive analysis. So this was not a way out.

Thus the traditional epistemologist finds himself forced to the conclusion that either there are reductive analyses in the various areas of knowledge, or else scepticism is true. And he can't find any reductive analyses.

III. THE MEANING OF A STATEMENT

I think that the solution to this difficulty in which the traditional epistemologist finds himself is to deny that the meaning of a statement must always be given by stating its truth-conditions. I shall now argue that there is another way of giving the meaning of a statement—by giving its justification-conditions. Or, to be more precise, the meaning of a statement is always uniquely determined by the justification-conditions of it and its denial.

A. Concepts

First let us talk about concepts rather than statements. I shall argue that the meaning, or logical identity, of a concept is uniquely determined by its justification-conditions, where now by "the justification-conditions of a concept" I mean the conditions under which one would be justified in ascribing the concept to something.

Before going any further, I think it is wise to stop and ask just what is meant here by the "meaning" or "logical identity" of a concept. I talk about "the logical identity" of a concept

because, in the case of concepts, their identity is a purely logical matter. If two concepts are identical, it is necessarily true that they are. I shall also talk about the meaning of a concept, but this is probably an abnormal use of "meaning." Although there is certainly historical precedence for this use of "meaning," we do not ordinarily talk about the meaning of a concept. But given a particular concept, we can talk about what it means for something to be an instance of that concept, and so we can understand talk about the meaning of a concept as elliptical for that.

Now let me try to establish my claim that the meaning or logical identity of a concept is uniquely determined by its justification-conditions. In order to do this I shall turn to the question of what is necessary before a person can be said to *have* the concept, or to have learned the concept. Given a concept, there is a set of conditions *C* which must be satisfied before a person can be said to have learned the concept. As different things are required in order for a person to have learned different concepts, the conditions *C* will be different for each concept. Thus a concept is uniquely determined by the set *C* of conditions which must be satisfied in order for a person to be said to have learned that concept. What I shall do here is argue that the conditions *C* are in turn uniquely determined by the justification-conditions of the concept, and thus those latter conditions also uniquely determine the concept.

Just what is necessary before we can truly say of a person that he has learned the concept of a certain kind of thing, such as "red thing" or "bird"? We frequently talk about a person "knowing what a bird is" rather than about his having the concept of a bird. But we must be careful with the locution "*S* knows what a bird is," because it can be used to mean two quite different things. There is a perfectly good sense in which a biologist knows more about what a bird is than does a layman, e.g., he knows that birds are warm blooded; that they have livers, etc. But this sort of knowledge about birds cannot be part of having the concept of a bird. All that we can require of a person in order for him to have a concept is "conceptual knowledge"—knowledge of things that are constitutive of the concept itself—and not knowledge of contingent facts. Someone must already have had the concept of a bird before these contingent facts could have been discovered. For example, if the biologist who first discovered that birds are warm blooded hadn't already known

what a bird was, he could not have told that it was a bird that he was examining.

It is the sense of "*S* knows what a bird is" in which the biologist had to know already what a bird was that is equivalent to "*S* has the concept of a bird." This just means "*S* knows a bird when he sees one." If *S* does not know a bird when he sees one, then he doesn't know what a bird is, regardless of the amount of theoretical knowledge he may have about birds. On the other hand, if *S* does know a bird when he sees one, even though he may not be able to tell us what makes a bird tick, then he knows what a bird is—he has the concept of a bird.

To say that *S* knows a bird when he sees one is just to say that he can pick birds out from other things—he can identify birds. Thus we can say that to learn the concept of a bird is just to learn how to identify birds. But now, this is still not as clear as we might desire. What is necessary in order for a person to know how to identify birds? There is perhaps a temptation to say that one must know what would count as making "*x* is a bird" true—one must know the truth-conditions for "*x* is a bird." But it doesn't take much reflection to see that this is not the case. Although a philosopher or a lexicographer *might* (although I doubt it) be able to construct a definition of "bird" which would give us such a set of truth-conditions, very few ordinary speakers of English would be able to do that, and we could not by any stretch of the imagination maintain that those ordinary speakers do not have the concept of a bird.

A person may very well have the concept of a bird without being able to give a definition, or give truth-conditions for "*x* is a bird." But one might suppose instead that in order for a person to know how to identify birds, although he need not be able to say what makes something a bird, he must nevertheless "do the identifying"—ascribe the concept "bird" to things—just when the truth-conditions are satisfied.

But if this is understood as requiring that the person never make a mistake, then clearly it is too stringent a requirement. Suppose we are teaching a child what a bird is. The simple fact that the child makes a mistake, thinking perhaps that a very cleverly constructed duck-hunter's decoy is a bird, is not sufficient to show that the child has not learned what a bird is or learned how to identify birds. The reason this would not count as showing that the child has not learned what a bird is is that he might have been perfectly justified in

thinking that the decoy was a bird, because it looked so much like one that anyone would have been justified in taking it to be one. If his ascription of the concept were thus justified, even though false, this would excuse his mistake from showing that he has not learned what a bird is.

Suppose then that the child ascribed the concept "bird" to things in cases where his ascription could not possibly be justified on grounds like the above. Would this count as showing that he has not learned the concept correctly? No, it wouldn't, because the child might know better than to ascribe the concept on such flimsy grounds, but do it anyway. He might not have been paying attention. Perhaps he was just careless. The simple fact that the child occasionally ascribes the concept "bird" to things unjustifiably would not in itself show that he has not learned the concept correctly.

But if in fact the child did not *know how* to ascribe the concept and its complement justifiably (i.e., he did not know how to determine justifiably whether something was a bird), this would show that he had not learned how to identify birds and so had not learned the concept.

Conversely, when the child has learned to judge justifiably whether a thing is a bird (i.e., he has learned to ascribe the concept and its complement to things justifiably), we are satisfied that he knows how to identify birds and so has got the concept right—he knows what a bird is.

Of course, when we say that the child must "know how" to ascribe the concept to things justifiably, this is knowledge in the sense of practical knowledge, rather than theoretical knowledge. The child must *know how* to ascribe the concept to things justifiably, but he need not *know what* is required for his ascription of it to be justifiable. Few people would be able to articulate the grounds on which they justifiably judge that something is a bird.

Thus far I have concerned myself only with the concept "bird," and I have concluded that to learn the concept "bird" is to learn how to ascribe justifiably the concept and its complement to things. The question that now arises concerns the extent to which this conclusion can be generalized to apply to other concepts. It seems clear that the conclusion can be generalized immediately to cover all concepts of *kinds* of things, such as the concept of a red thing, the concept of a four-legged thing, the concept of a cat; etc. It can also be generalized to cover the concepts of relations between things, such as "brother of" or "taller

than," because those can be thought of as concepts of kinds of ordered pairs (or more generally, ordered n -tuples). It is not clear to me whether all concepts can be thought of as concepts of kinds of things, and so brought under the above discussion, because it is not clear to me just what the boundaries of our concept of a concept are. But it at least seems evident that a great number of interesting concepts can be included in our discussion. In general, whenever we have a concept that can be thought of as the concept of a kind of thing, then we can say that the meaning or identity of that concept is uniquely determined by the conditions under which we would be justified in ascribing it or its complement to something.

It seems then that to learn a new concept is to learn how to judge justifiably what things are instances of that concept. But this is just to say that to learn a concept is to learn how to ascribe it and its complement to things when the justification-conditions are satisfied. And thus the identity of a concept is uniquely determined by the justification-conditions of it and its complement. The justification-conditions are definitive, and thus, essentially, constitutive of the concept.

B. Statements

Having discussed the identity of a concept, we can now turn to the identity of a statement. In analogy to my discussion of concepts, I shall argue that the identity (or meaning) of a statement is determined by the justification-conditions of it and its denial.

A statement states that something is the case. E.g., the statement that Smith is a bachelor states that Smith is a bachelor. This is a truism. Statements are differentiated by what they state to be the case, or by the kind of state of affairs in which they would be true. For example, the statement that Smith is a bachelor would be true in any state of affairs which included Smith's being a bachelor, and so by picking out this particular kind of state of affairs we can uniquely specify the statement: that Smith is a bachelor by saying that it is that statement which would be true in just this kind of state of affairs. Thus the identity of a statement is uniquely determined by the kind of state of affairs in which it would be true.

Kinds of things are correlative with concepts of

kinds of things. That is, two different *concepts* of kinds of things (e.g., the concept of a blue thing and the concept of a red thing) determine two different *kinds* of things, and conversely, given any two different kinds of things, the concepts of those two different kinds of things are distinct (e.g., the concept of a blue thing is distinct from the concept of a red thing). Therefore, as the identity of a statement is uniquely determined by the kind of state of affairs in which it would be true, it is also uniquely determined by the *concept* of the kind of state of affairs in which it would be true.

We have already seen that a concept is uniquely determined by the justification-conditions of it and its complement. Thus the concept of a kind of state of affairs is uniquely determined by the conditions under which we would be justified in judging that the present state of affairs is or is not of that kind. Therefore the concept of the kind of state of affairs in which the statement that *P* is true is uniquely determined by the conditions under which we would be justified in judging that the present state of affairs is or is not one in which the statement that *P* is true. But these latter conditions are just the conditions under which we would be justified in judging that the statement that *P* is or is not true, i.e., the justification-conditions of the statement that *P* and the statement that it is not the case that *P*. Thus we have proven what we set out to prove: The identity of a statement is uniquely determined by the justification-conditions of it and its denial.⁴

IV. THE DEMISE OF TRADITIONAL EPISTEMOLOGY

Now let us return to the difficulty in which the traditional epistemologist finds himself. The traditional epistemologist is trying to justify sources of knowledge. The problem of justifying sources of knowledge is not to justify the source of knowledge for some particular statement, but to justify the source for all statements in certain classes, such as the class of material object statements, or the class of ethical statements. What we must do is prove, on the basis of the meanings of these statements, that these sources of knowledge do justify our beliefs in these statements. A problem arose when

⁴ This principle bears certain obvious similarities to the Verification Principle of the Logical Positivists, and in fact, I would suggest that it is just this similarity that has always made the Verification Principle seem plausible even though it could not be stated in an unobjectionable manner.

it was assumed that this meant that we must prove this on the basis of the truth-conditions of the statements, because that required reductive analyses, and reductive analyses don't seem to be possible. But I have just argued that there is another way of giving the meaning of a statement—by giving its justification-conditions. For those classes of statements for which reductive analyses are impossible (and these are most of the philosophically interesting ones), we can still give their meanings in terms of their justification-conditions. Thus, for *these* statements, the problem reduces to one of proving on the basis of the justification-conditions of these statements that the sources of knowledge on which they are based do justify our beliefs in them. But stating those justification-conditions *just is* stating what is a source of knowledge for these statements. Thus the problem of proving on the basis of the meanings of these statements that their sources of knowledge do justify them has a trivial (tautological) solution, because their meanings are given by those sources of knowledge in the first place. Where reductive analyses are impossible, the problem of justifying the sources of knowledge becomes uninteresting because of the triviality of its solution. But, and this is important, it *does* have a solution, even if only a trivial one, and there is no danger of scepticism rearing its ugly head.

The cases that have traditionally interested epistemologists are just the cases in which reductive analyses seem to be impossible. And that is in fact just the reason those cases seemed interesting. They were interesting because the traditional tools of the philosopher—truth-conditions and reductive analyses—seemed to be incapable of handling them. So my conclusion is that the traditional problem of justifying different sources of knowledge turns out in general to be a bogus problem. It usually can only be done in a trivial manner. Does this mean then that epistemology as a discipline is uninteresting and is better ignored? No, it does not. Let us consider again what I called “the fundamental problem of epistemology.” That was the problem of answering questions of the form “How do you know that *P*?” These are requests for explanations concerning what justifies us in the different kinds of knowledge claims that we make. These requests can be understood in either of two ways: (1) What are in fact good grounds for believing that *P*? (What are the justification-conditions for the statement that *P*?); (2) How can we justify basing our belief that *P* on the kind of

grounds that we ordinarily take to be good grounds? The best way to see the difference between these two questions is to see what it would take to answer them. (1) can be answered just by stating the justification-conditions of the statement. To do this we might rely simply upon our intuitions concerning when we would take the statement to be justified (just as before philosophers were content to rely upon their intuitions to discover truth-conditions). (2) on the other hand cannot be answered simply by stating the justification-conditions. (2) requires you to *justify* those justification-conditions, which implies you must derive them from something deeper. But, as we have just seen, there isn't generally anything any deeper.

The difference between questions of type (1) and questions of type (2) can be illustrated by returning to the problem with which this paper began. How do I know that I am not wired into a computer which manufactures all of my experiences in such a way that they are independent of the way the world really is? If this is interpreted as a question of type (2), then the answer is trivial. The answer is that the experiences that I in fact have are sufficient to justify me in believing that many things are true of the world around me. That these experiences are sufficient is constitutive of the concept of the material world. This is simply part of the justification-conditions of statements about material objects. But some of the statements about the material world that I am thus justified in believing are statements about my present surroundings, and these statements are such as to preclude the possibility of my being wired into an experimental apparatus of the sort envisaged. This then is how I know that I am not in the position of the subject of the fictitious experiment. But, clearly, this answer is not of much interest because of its triviality. On the other hand, if the question is interpreted as of type (1), then its answer is of considerable interest. In order to answer it, one would have to get clear on just what are the justification-conditions of statements about material objects, and that is obviously a matter of great complexity.

I have argued that philosophers shouldn't concern themselves with questions of type (2) because they have only trivial solutions. But questions of type (1) remain as interesting problems. The problem is to state in a precise manner just what is involved in justifying different kinds of knowledge claims. For example, we want to know exactly what the relationship is between perception

VI. ON POINTS OF VIEW

JON MOLINE

WHAT is the point of a point of view? Much is made of the notion.¹ Popper speaks of a "logical point of view,"² Toulmin of "new points of view" in the sciences,³ and pundits of a "military point of view." Often the expression "point of view" is not a mere expression, but a piece of somewhat technical terminology. It is often a load-bearing member in arguments, and the understanding of some of these arguments will be found to turn on the correct account of what it is to take a point of view. One may hear it said, for example, that from a scientific point of view, Lysenko's "data" were irrelevant to the genetic theory he wished to refute, that in order to understand perplexing actions, one must understand the point of view of the agent, and that disputes between nations taking a narrowly nationalistic point of view can be settled by inducing them to take a larger point of view. Such uses of the expression "point of view" move one to ask "What is it to take a point of view?"

An adequate answer would make clearer what one is asked to do when it is suggested that one "take a larger point of view." It would also help to resolve certain problems and dilemmas people pose for themselves and for others. In some peoples' minds there lingers the suspicion that if on a given issue one takes the point of view of self-interest, one cannot take the moral point of view.

A corollary of this suspicion suggests that if on a given issue one takes the moral point of view, one cannot take the point of view of self-interest. Most of those who harbor such suspicions would like to avoid both horns of their dilemma, and suspect that there is at least sometimes a way between them, but this way is not clear. One reason it is not clear is that no adequate answer has been given to the question "What is it to take a point of view?"

I

The literature on points of view is scant. Kurt Baier is one of the few philosophers ever to make any remarks suggestive of a general account of what it is to take a point of view. In *The Moral Point of View*, he submits that

Answers to practical questions are arrived at by reference to a point of view, which may be defined by a principle. When we adopt a point of view, we adopt its defining principle. To look at practical problems from that point of view is to be prepared to answer practical questions of the form "What shall I do" "What shall be done" by reference to its defining principle.⁴

Baier almost certainly did not intend these few remarks to stand as a complete, general answer to the question "What is it to take a point of view?" But it will be instructive to treat them as such an answer, and to note where difficulties arise. First, it should be noticed that it does not seem to be the case that every point of view can be defined by "principles," far less by a single "principle."⁵ Even Baier himself does not attempt to define the moral point of view in terms of a single "principle." And what "principle" could be said to define the Negro point of view? Journalists and Negroes speak of such a point of view, and others attempt to understand it; yet no one, to my knowledge, would attempt to define it in terms of "principles."

Baier suggests that to adopt a point of view is to be prepared to answer practical questions by reference to a "principle." But again, if this suggestion were to apply to all points of view, difficulties would arise. Some have adopted or attempted to adopt a Negro point of view by darkening their

¹ An infallible sign of the recent popularity of the notion is provided by the Beatles, who, in their song "Nowhere Man," rebuke the person who "doesn't have a point of view."

² Karl Popper, *The Logic of Scientific Discovery* (New York, 1959), p. 27. Cf. The title of W. V. Quine's *From a Logical Point of View* (New York, 1963).

³ Stephen Toulmin, *The Philosophy of Science* (New York, 1960), p. 23.

⁴ Kurt Baier, *The Moral Point of View* (Ithaca, 1958), p. 184.

⁵ *Ibid.*, pp. 191, 195, 207. Baier's characterization of the moral point of view and of what it is to adopt it is actually very complex, and apparently involves "principles" in only a peripheral way.

skins and attempting to have certain experiences, to feel the brunt of certain attitudes, and the like, not by adopting "principles."⁶ Learning and using principles may be involved in adopting some points of view, but not in all; and much more is involved. If one wishes to adopt a physicist's point of view, one must learn a great many facts, techniques, and equations, few of which are readily characterizable as "principles." If one wishes to view things from a schizophrenic's point of view, the only convenient way is to take a drug, perhaps LSD. It would be ludicrous to suppose that there are principles learned and used here. Hence, if we take Baier's remarks on what it is to adopt a point of view as a general account, it is an inadequate one.

As a general account of what one who has a point of view does, Baier's remarks are also inadequate. He submits that one attempts to answer practical questions. This is, of course, true of one who takes the moral point of view. But it is not clear that the schizophrenic tries to answer any questions, although he may be said to have a point of view. And questions other than the practical ones Baier emphasizes are answered by those taking certain points of view. It is difficult to think of many of the questions dealt with by those taking a chess-player's point of view as practical ones.

Baier's brief remarks, therefore, tantalize rather than enlighten us on the question "What is it to take a point of view?" Much more must be done if we are to answer the question adequately.

II

The expression "point of view" and the terms "adopt" or "take" have of course various senses. I shall (A) label and sketch these, (B) examine the sorts of claims we make by using the expressions in these senses, and (C) trace the scope within which we use them.

(A) It is useful to distinguish at least two senses of the expression "point of view." The first is the sense in which "this point of view" is taken to mean nothing more than "this spatial location," regarded literally as a vantage point of the sort a photographer might seek. This prosaic first sense has philosophic significance only in that the second is probably an extension of it and has interesting

similarities to it. From a vantage point one is able to notice certain things but not others, and one can perform certain tasks but not others. Some vantage points, such as the very top of an extremely narrow peak, can be taken by only one person at a time. Other points of view, like some peaks, are sufficiently "broad" that many persons could be said to be looking at things from the same point of view or perspective at the same time. Whether or not two people have the same vantage point will depend on elevation, visual obstructions, distance from objects being viewed, contours of terrain, and the like. These features of points of view in the location sense have counterparts in the sense to be considered next.

The second sense I shall call the philosophic sense, although it is by no means confined to philosophy. In this sense, a point of view is not a place from which one views things and events but a way of viewing them. Like points of view in the location sense, points of view in the philosophic sense may be such that they can be taken literally by only one person at a time, or they may be literally sharable by many people. The President's point of view may be unique, but the military point of view is quite widely and literally taken. The military point of view is not taken, however, merely by standing behind the desk of the Chairman of the Joint Chiefs of Staff, nor could the President's point of view be taken merely by standing in a certain place. When pundits speak of the President's point of view they are not speaking of what he can see from the windows of the Oval Office.

Obviously, this second sense is the philosophically important one. Hume, for example, appears to have been using the expression in this sense when he spoke of a "point of view common . . . with others" taken by one using moral language.⁷

There are also various senses in which one can "adopt" or "take" a point of view. Not all of these involve one's acting overtly.⁸ The actions of a store detective in a large department store proceed from his having adopted a detective's point of view, yet he is justified in "adopting" in a sense a shoplifter's point of view. He does so not in overt action by stealing from his employer, but in thought, as a tactically useful intellectual exercise designed to uncover weaknesses in the store's defenses against thieves. He asks, "Now if I were a shoplifter,

⁶ John Howard Griffin, *Black Like Me* (Cambridge, 1960), pp. 1-3, pp. 11-14.

⁷ David Hume, *An Enquiry Concerning the Principles of Morals* (London, 1751), Section IX, Part I.

⁸ Baier seems to have overlooked this in his preoccupation with practical, moral questions.

how . . . ?" He merely attempts to understand the shoplifter's way of looking at conditions in a store. This intellectual exercise no doubt affects his actions, but only indirectly, and not in the way that his adopting a store-detective's point of view does. His adopting a detective's point of view determines his ends; his adopting a shoplifter's point of view at most renders more subtle the means he chooses to achieve these ends. He *is*, *acts*, and *thinks* as a detective. He merely thinks as a shoplifter. Hence there are at least two senses in which one may "adopt" a point of view. One may overtly take and act on a point of view *P*, or one may merely try to understand *P*.

The distinction between these two senses in which one can adopt a point of view is reflected in the purposes of literature which might be said to characterize certain points of view. The *Soldier's Guide* is intended to tell one how to go about soldiering; biographies of Napoleon are not intended to be convenient guides to Napoleonizing.

Some points of view elude our understanding or adoption even in the second and weaker sense. The most sensitive and skillful investigator may be unable to understand the point of view of a person who is uncommunicative, inarticulate, or unable to provide him with the requisite information.

One may know a good deal about a given point of view without "adopting" it in either of the above senses.⁹ One knows that failing to pay one's income tax is a legal matter, that it is sometimes punishable by imprisonment, and the like, and yet one does not have to adopt the legal point of view in order to realize this. One may become aware of it merely by noting that those having certain legal roles treat such failure as falling within their professional domain. If one had no idea whatever of the process by which they arrive at the judgments they do regarding such acts, one could not simulate their doing so even in thought, and therefore could not in either of the above senses adopt the legal point of view.

(B) It should be noticed that there are at least three principal sorts of claims made by using the expression "point of view" in the philosophic sense, and that these claims may involve a suggestion that one merely try to *understand* a given point of view, a characterization of such a point of view, or a critical assessment of the point of view, and hence

of *taking* it. The first sort of claim is made in explanatory contexts. One hears it said, for example, that "If only you would try to understand the point of view De Gaulle is taking, you would not find his decisions so incomprehensible—he is taking a purely nationalistic point of view." I shall refer to comments of this type as "comprehension claims".

Claims to the effect that a certain consideration is relevant or irrelevant from some point of view or other constitute a second class. It is urged, for example, that psychological matters are irrelevant from a philosophic point of view. Claims of this sort I shall call "irrelevance claims."

One sometimes encounters rather critical claims of a third sort made by means of the expression "point of view," as in the comment "Senator Yawn takes a narrow, economic point of view on all proposed legislation." Two of the most common critical qualifiers used of points of view are, significantly, "narrow" and "restricted." Points of view are commended, however, as "larger," "broader," and "wider." I shall refer to such claims as "size claims."¹⁰

(C) Consider next the scope within which one may use the expression "point of view." This scope appears to be restricted by considerations of personality and role, although the restriction is in neither instance a tight one. We may shed light upon the relevance of considerations of personality by considering groups of examples. We speak freely of all the following with no sense of oddness:

- I. a physicist's point of view
an administrative point of view
a parental point of view
the French Government's point of view

All the following, by contrast, are very odd:

- II. a coffee-break's point of view
a cello's point of view
the point of view of the square root of 2
a compliment's point of view
a rock's point of view

A third group of examples falls between the first and second in oddness. We say the following with a bit of strain, if at all, but we might nevertheless find some occasion for saying them:

⁹ I am indebted to Professor T. E. Patton for this point.

¹⁰ These claims are discussed further in Pt. IV below.

- III. a dog's point of view
 a computer's point of view
 a clam's point of view
 a worm's point of view

It is often profitable to ask the basis for such differences in degree of oddness as we discern in the three groups above. The answer often sheds light on philosophically troublesome concepts. Here, a clue as to how to approach the answer to the question "What is it to take a point of view?" is readily forthcoming from the comparison: We might say, in a moment of affectionate anthropomorphism, "From our dog's point of view, he is master of the house—we are merely his servants."¹¹ Clearly this is an extension of the usual scope within which we use the expression "point of view," an extension which depends upon our granting to the dog, by an affectionate sort of courtesy, the same status held by an imperious person. Why are we prepared to make such an extension in the case of a dog, but not in the case of a clam?

The answer seems to be that we are able to regard the dog as we would a person or "quasi-person" and that this is necessary in order for the use of the expression "point of view" to seem a sensible one. The dog may reasonably be said to have a personality. A clam may not. One clue, then, to the appropriateness or inappropriateness of using the expression "point of view" is that such use seems to depend upon how easily one can view the subject said to have the point of view as having a personality.

Another clue is that there often seems to be some sort of loose connection between taking a point of view and having a position in the metaphorical sense in which "having a position x " means "filling role x ." All the entities, groups, or persons in group (I) are associated with certain roles; roles which, of course, only persons or groups of persons can fill. Administrators and parents, for example, have fairly distinctive roles and associated with these roles are distinctive points of view. Even the dog in the example above was viewed as playing a role—master of the house.

Those entities in the second group have neither roles nor personalities, although all of them perhaps have functions. The Lord Aylesford Stradivari has a function, but no role. Starker has the role. I am not concerned with giving a full analysis of roles,

however, only with using the apparent loose connection between roles and points of view as a clue to what it is to take a point of view; but the above distinction between roles and functions is not difficult to make out in the intuitive way needed here—roles are to functions as actors are to props. Non-inanimate objects have roles.¹² Non-human animals are spoken of as having them only by courtesy, a courtesy involving an extension of the ordinary domain within which we apply the expression. This courtesy seems a sensible one only if the subject said to have a role can be viewed without a great deal of strain as having a personality.

But we are able to link points of view and roles only loosely. Raskolnikoy has no role corresponding to his particular point of view. In order to understand his point of view, we may need to understand the various roles he has, but this alone will not give us an understanding of his point of view. Similarly, there is a fairly large number of paranoids and it makes sense to speak of a paranoid's point of view, but it eludes understanding to speak of a paranoid's role.

The connection between roles and points of view is thus a fairly loose one. Yet it can shed light on what it is to take a point of view. Consider what a role is. The theater provided the term, which originally meant the roll or script upon which an actor's part was written. The term still suggests what it is given to a certain character to say or do, what is appropriate for him to feel, express interest in, try to accomplish, or assume while playing that role or remaining "in character."

The sense of "role" in which roles are often associated with points of view is a slightly more figurative one suggesting not a person simulating another functioning in some capacity, but one really functioning in a certain capacity or position. Both having a point of view and filling a role seem to involve acting or behaving in certain ways. Hence our consideration of the connection between roles and points of view suggests that it will be helpful to replace our original question "What is it to take a point of view?" with the question "What does one taking a point of view do?"

III

The answer to this question is in part suggested implicitly above, but examples will be of use in

¹¹ Apparently this is an implicit comprehension claim. The person making it intends to shed light—most of it ironic—on the dog's behavior.

¹² I shall ignore here the possible, though controversial, exception of computers.

clarifying it and in bringing to light the rest of the answer.

Consider the role and associated point of view of a businessman. It might be said that a businessman's role in a community is to provide goods and services in a manner satisfactory or at least not radically unsatisfactory to the community while providing employment for himself and perhaps for others. We all know fairly well what his adopting the businessman's point of view involves, and that he probably assumes it in order to facilitate his filling that role. It is typical of points of view that they are assumed in order to give one guidance as to how best to fill some role or another. It is more helpful, therefore, to ask, "What does one taking a businessman's point of view do?" than "What is it to take a businessman's point of view?"

We expect one taking a businessman's point of view to regard a fairly limited number of factors as directly and importantly relevant to his decisions from this point of view. For any given action x , considered from a businessman's point of view, we expect at least these questions to be asked:

- (1) Is x profitable in any immediate sense, and if so, to what extent?
- (2) Is x legal?
- (3) Will x satisfy or dissatisfy customers or otherwise affect profits in the long run?

Presumably a businessman will not, *qua* businessman, ask any of the following except insofar as they bear indirectly upon the answers to 1-3:

- (4) Is x morally right? (But he may ask, "Will my customers be morally offended by x ?"")
- (5) Is x aesthetically pleasing? (But he may ask, "Will my customers be aesthetically offended by x ?" or "Will potential customers be aesthetically attracted by x ?"")
- (6) Is x conducive to the advancement of learning? (But he may ask, "Is x likely to gain for me technical information of use in answering 1-3 for some y ?"")

Taking a point of view, then, is like picking up and looking through a lens constructed for a particular purpose and having a particular focal length and

field of vision. Some objects will be in focus (relevant) and others will simply be excluded from view (irrelevant). Taking a point of view is like adopting a metric standard—it enables one to accomplish certain tasks by referring to certain authoritative marks. It is also like making assumptions in an argument—one can then generate conclusions.

In general, we expect one who has taken or adopted a certain point of view to display at least the following tendencies:

- (a) a tendency to have and pursue certain specifiable interests and aims,
- (b) a tendency to use only certain criteria for the evaluation of actions as conducive to achieving these aims or not,
- (c) a tendency to regard other interests and criteria as largely irrelevant, or at most relevant only in an indirect and limited way,
- (d) a tendency to make certain factual assumptions as, e.g., the paranoid's assumption that all facts can be explained by reference to a grand conspiracy, or De Gaulle's assumption that the U.S. and Great Britain are not to be trusted.¹³

We would also expect the person taking a certain point of view to display

- (e) a tendency to agree with or share the interests, goals, criteria, assumptions, and judgments of relevance and irrelevance characteristic of others taking the same point of view.¹⁴

Adopting or taking the same point of view will in fact consist in adopting and sharing the same assumptions, criteria, interests, goals, and judgments of relevance and irrelevance. To say that two people have the same point of view is to say that they tend to share these.

It should be noticed that one can be said to have (though not to have taken or adopted) a certain point of view if one merely appears to have tendencies such as (a)—(d). Paranoids, for example, may or may not actually share certain interests, assumptions, criteria, goals, and judgments of relevance and irrelevance. We can sensibly speak of a paranoid's

¹³ Tendency (d) was pointed out to me by R. Harré.

¹⁴ This claim requires qualification, for it holds only when such interests, goals, assumptions, and judgments of relevance and irrelevance are specified in general terms. Those taking a military point of view would all agree that one of their primary goals is victory, but this goal cannot be further specified as, e.g., "U. S. victory," since some of those taking the military point of view desire victories incompatible with this.

point of view not because we know that paranoids share these, but because they act as if they did. But if we adopted the general account suggested by Baier's remark that adopting a point of view is defined by using "principles," we would render unintelligible this and similar uses of the expression "point of view" and obscure the function of the expression. People may be said to have points of view (e.g., the paranoid's) which they cannot sensibly be said to have adopted or taken. This may well apply to the moral point of view itself; surely not all those who are said to look at things from a moral point of view can be said to have *taken* or *adopted* that point of view at some time or other; and it seems unlikely that in saying their point of view is a moral point of view we always intend to say that certain mental events are going on in them. The expression "point of view" does not appear to be used simply to describe certain mental events, actual or ideal. It appears to be used in the rational reconstruction, interpretation, and criticism of both rational and irrational tendencies in behavior.

IV

Having provided this answer to the question "What does one taking a point of view do?" or, better, "What is involved in taking a point of view?", I should now like to examine and elaborate upon the answer.

This account of what is involved in taking a point of view can be tested first by asking whether it sheds any light on the dilemmas posed in the introductory section. If, for example, one takes a point of view P_1 on a certain issue, does it follow that one cannot at the same time on the same issue take another point of view, P_2 ? Or does it necessarily follow that whatever is sanctioned by a point of view P_1 is either not sanctioned or forbidden by any point of view P_2 ? Intuitively, we know that certain points of view overlap, and that the answer to these questions is sometimes "No." But why (and hence, when) it is "No" can now be clarified. Consider the following case: Suppose that one takes a point of view P_1 in a case we shall call C_1 , and that taking this point of view consists in assuming interests I_1, I_2, \dots, I_n , using criteria R_1, R_2, \dots, R_n , and making factual assumptions F_1, F_2, \dots, F_n . Now in case C_1 it is unlikely that all of these interests, criteria, and assumptions will be relevant; perhaps the case gives opportunity for the satisfaction of only one of these interests, say I_j . But I_j may be an interest proper to

another point of view, P_2 . Hence if one acts in the way sanctioned by point of view P_1 in case C_1 one may be said to have acted in a way consistent with (and perhaps demanded by) point of view P_2 . If one was aware of this at the time, and would not have done what was demanded by point of view P_1 unless it was also sanctioned by point of view P_2 , there is reason to say that one took the two points of view P_1 and P_2 on the same issue.

Similar remarks apply for the relevant criteria and factual assumptions. Points of view may overlap in their aims, their criteria, their assumptions, or the courses of action they demand, sanction, or forbid. It might be the case that the moral point of view and the point of view of self-interest sanction the same course of action in a given case. To show that they did, one would have to review in the light of the facts of the case the relevant interests, criteria, factual assumptions, and courses of action demanded, sanctioned, or forbidden in order to show in precisely what respects these were the same for the two points of view. If two points of view are incompatible, this could again be shown only by reviewing these factors.

An additional test of the account can be provided by trying to discern how it clarifies the three sorts of claims mentioned above, comprehension, irrelevance, and size claims.

One sometimes hears it said, "You would not be so mystified by the Senator's policies if you understood the point of view he is taking—it is that of political expediency alone." One who says this is attempting to explain the Senator's behavior. And in the light of the above account of what is involved in taking a point of view, it is easy to see how this is implicitly an explanation. Citing assumptions, interests, goals, criteria, and restrictions on considerations thought relevant are common ways of explaining human behavior. And if one knows these factors for a certain person, one may even be able to predict future behavior. No new problems, then, seem to be raised by this account in connection with comprehension claims.

The account does, however, raise at least one pressing question with regard to irrelevance claims, viz., what is it that makes a consideration relevant from a particular point of view? If the account is to clarify such claims, this question must be answered.

The most plausible answer seems to be that a consideration C is relevant from a point of view P to the extent that ignoring C prevents, interferes with, or fails to take advantage of an opportunity

to facilitate the accomplishment of the aims which are characteristic of *P*.

This seems a plausible criterion of relevance, and it is difficult to envision one which is any more useful, yet one should not give way to the illusion that it will support very many bold irrelevance claims utilizing the notion of a point of view. Practically any consideration might conceivably be relevant in a minimal way from almost any point of view on this criterion. And this is not implausible. Neurologists, e.g., continually inform psychiatrists of physiological factors relevant to the accomplishment of the psychiatrist's tasks, factors which earlier were either unknown or dismissed as virtually irrelevant.

It might, however, be suggested that a general criterion of relevance is fruitless, and the suggestion might be supported by the following argument: If the aesthetic results of some action of a businessman (say, the erection of a remarkably ugly sign on the front of his shop) cost him one customer, he may dismiss this as irrelevant from a business point of view. If his business is a small one, however, and the action costs him 500 customers, he will certainly regard this as relevant. If the loss of one customer is irrelevant and the loss of 500 is relevant, then at precisely what point do the aesthetic results of actions become relevant from a business point of view? An argument of the sorites type seems to show that it is impossible to fix this point with precision, and even if it were possible, the point would plainly be relative to the individual businessman, and hence, like Aristotle's mean, would be of little or no general use in guiding action. Macy's can well afford to lose 500 customers because it has offended them aesthetically, but the corner druggist cannot. But general remarks on criteria of relevance and irrelevance will help neither Macy's nor the corner druggist.

However, this argument is not compelling. One might as poorly argue that general remarks upon shirt-fitting are fruitless because shirt sizes are relative to individuals. General remarks upon criteria of relevance are no more vitiated because of their applying in different ways to different individuals than general remarks upon shirt-fitting are vitiated because shirt-sizes are relative to individuals.

But the above criterion of relevance allows one to meet the argument in a better way. "Irrelevant" in the claim "the loss of one customer is irrelevant from a business point of view" means not "totally irrelevant" but "negligible." Even Macy's would

concede that the loss of one customer is a bad and relevant thing from a business point of view, nevertheless the loss might be negligible for Macy's because it is the lowest possible and hence unavoidable. Therefore, saying that the loss of one customer is irrelevant to a particular business either (a) unpacks to a supportable claim that given the goals, criteria and assumptions of a business point of view, reconsidering it would not or could not in any way affect the decision which gives rise to it, or (b) expresses a stubborn and groundless refusal to reconsider evidence.

Thus, the answer to the question above, "What makes a consideration relevant from a particular point of view?", is that a consideration is relevant in a particular case if ignoring it avoidably interferes with the accomplishment of the aims characteristic of the point of view in question. A consideration is irrelevant if considering it from this point of view, however carefully, would have no effect or negligible effect. It is wholly irrelevant in the unlikely but nevertheless possible event that it is entirely inconceivable how it could ever affect decisions from this point of view.

With the addition of this criterion of relevance, the above account of what it is to take a point of view seems to render irrelevance claims more intelligible.

Size claims are more difficult to deal with in giving an account of what it is to take a point of view, however, for they may have a number of meanings. Suppose one makes an adverse size claim about Secretary M——'s point of view in diplomacy by saying "Secretary M——'s point of view in diplomacy is much too narrow." This claim may have any one or perhaps all of a number of meanings. The above account suggests that it might possibly mean:

- (a) that Secretary M——'s interests are too narrow,
- (b) that his criteria for deciding what lines of diplomacy to pursue are unduly restricted ones,
- (c) that he persists in dismissing wider interests and criteria having greater scope and complexity,
- (d) that his lines of diplomacy are based upon too small a body of relevant facts.

Until the claim is specified in one or more of the above ways, its warrantability cannot reasonably be assessed.

But there are serious ethical pitfalls to be avoided in interpreting size claims. For example, one making size claims might be bothered by such questions as "How can one best decide which among alternative points of view one is justified in taking?" He might answer, "By noticing which of relevant alternative points of view is 'largest' and by adopting that one." He might then interpret the "size" of a point of view in terms of the number of people having an interest in the aims of the point of view. Then he might insist that by "interest" he does not mean psychological apportionment of attention, but rather "welfare." It is fairly clear how such an adherent of size claims might thereby purchase many of the difficulties inherent in utilitarianism for what may be no more than the advantage of rephrasing old ethical questions in a new way. But this problem is generated not by the above account of what it is to take a point of view but by one particular ethical interpretation of the basis for size claims. And other interpretations are possible. I am not concerned here with giving any

such interpretation. To do this would involve formulating an ethical theory, whereas my task in this paper is the general understanding of the notion of a point of view and the claims in which it appears, not the support of such claims. And, as we have seen, the account of what it is to take a point of view in terms of tendencies (a)–(d) seems to render such claims more intelligible.

In summary, our original question "What is it to take a point of view?" led to questions concerning what one does in taking a point of view, three types of claims one can make by using the expression "point of view," various senses in which one can adopt a point of view, and ways in which points of view overlap. Answers were proposed to all of these questions and to others arising out of them, with the exception of certain ethical questions provoked by size claims.¹⁵

One final question must also remain unanswered here, viz., "Is this account of points of view circular because it has proceeded from a philosophic point of view?"

The University of Wisconsin

Received May 15, 1967

¹⁵ I plan to deal with these ethical questions on another occasion.

VII. EVALUATION

JON WHEATLEY

IN this paper I wish to make two axiological points, one positive and the other negative. They are, in a sense, rather small points; but they play havoc with a good deal of the work in ethics we have had in this century. I shall start by noticing some of the ways we do evaluate things and people and actions and events every day. For the present, I shall avoid considering examples where we use the old philosophical paradigms of evaluation; that is, I shall not consider utterances containing "good," "right," and "beautiful" at first. I do this in the hope that we can get clear about at least some of what is involved in evaluating before getting into the knotty areas where we cannot see the functions of the idioms for the fashions of the philosophers.

2: Consider the following examples of evaluations:

- (1) I take you to an art gallery and draw your attention to a piece of modernistic sculpture. You say, "I don't like it."
- (2) The same situation. You say, "It makes me feel ill."
- (3) The same situation. You say, "Ough!" or "Ouch!"
- (4) The same situation. You say, "Wonderful!" or "Awful!"
- (5) I ask about Smith. "He's an old goat," you say.
- (6) I ask you what you object to about Jones. You say, "The man's a liar."
- (7) The art gallery situation again. You say, "It looks like a fugitive from a junkyard."
- (8) The same situation. You say, "It's sublime!"

It is in these and similar ways that we do pass judgment on, or evaluate, things, situations, people, and events every day.

One point is worth noticing immediately, though it is not argued fully here. Evaluation is not, in general and for the most part, a function of sentences, still less of predicates. Consider the list of

examples above and look only at what appears between the quotation marks, i.e., ignore the rudimentary context I have given these utterances. It is clear that the majority of these sentences could be uttered in circumstances and in ways which would remove all, or most of, their evaluative force. Even "Wonderful!" can be uttered sarcastically (though it then remains evaluative). If we ignore the possibilities in sarcasm, then "Wonderful!" cannot help but be a positive evaluative exclamation; but this is the only really clear case. "It's sublime!" also looks like a good candidate but there are difficulties about it. The sentence "It's sublime!"; appearing as a lone utterance, or such that its utterance constitutes the only part of an utterance relevant to evaluation, and when it is not used sarcastically or ironically, can perhaps only be positively evaluative. However, *the form of words*, "It's sublime!" need not be: consider, "It's sublime, I'm told," and even more, "It's sublime, I suppose but . . ." We can sum this point up by saying that, for the most part at least, it is acts (mostly speech acts¹) not idioms which evaluate. I shall return to this point periodically throughout the paper.

3. Example (1) is a straight statement of attitude or reaction or similar. To utter such a sentence appropriately is an easy way to evaluate and, of course, a perfectly good one. It is not, however, a very *powerful* way to evaluate. It is a statement of personal attitude or reaction, but like all such statements invites replies of the type, "You could learn to like it" or "I don't care what you feel, I do like it." It is for this reason that the position that every evaluation is a disguised statement of attitude is so grossly implausible; that is, because statements of attitude or reaction invite the replies they do, because they fit into discourse the ways they do; they are clearly quite unlike our stronger, more positive, evaluations.

Examples (2) and (3) respectively; state and express reactions or something very close to that; examples (3) and (4) could be said to express

¹ "Speech act" is Austin's term, though used in ways other than his by some writers. I shall use it as he used it, i.e., to mean the utterance plus the total situation in which it is uttered. As will be obvious to those who know his work, my debt to Austin in this paper is very large.

emotion not, indeed, in the ordinary sense of "express emotion" but in the sense made familiar by the Emotivists. But there is a point about these examples worth noticing which the Emotivists always, I think, overlooked. Such speech acts as (2), (3), and (4) work best as evaluations where the emotions, reactions, etc., involved are not, in fact, felt; certainly they need not necessarily be felt. To see this, notice the following: Suppose I react to something with an involuntary grunt of disgust. It is dubious if I have thereby necessarily evaluated what led me to grunt. *Because* it disgusts me, I may evaluate it negatively, but the actual grunt surely does not, of itself, evaluate. I might give an involuntary expression of disgust when watching, say, an abdominal operation; but I might also evaluate the operation, and even the business of watching it, highly. However if, not feeling particularly disgusted, I make an utterance whose primary function is to express disgust then, most likely, I am evaluating. Thus when we evaluate in the ways illustrated in (2), (3), and (4), the evaluation is susceptible of an analysis which depends on the primary use of these sorts of idioms to express emotion, state or express visceral reactions, and so on.

Let me put these points more systematically. There are a number of idioms in the language whose obvious and literal function is to express or state attitudes, emotions, reactions, and so on. These idioms can be, and often are, used to evaluate. But we do not, by and large, use them for careful, consistent, important evaluation. That is, though we use this general technique relatively frequently, we do not use it, as a rule, when we really care about what we are evaluating. To say just "Ough!" about a murderer is, we feel, inadequate. This is not surprising. This sort of evaluation is on a logical par with the school boy who, there being no genuine smell, holds his nose.

I call the idioms in which we express such evaluations "semi-evaluatives": "semi" because they are weak and inadequate for important matters, and "evaluatives" to remind us that it is acts, not idioms, which constitute evaluations.

4. There are a great many ways to evaluate, a great many idioms and types of sentences which can be turned to evaluative use. Below I discuss a reasonable cross-section of the ways we do in fact evaluate.

(A) One way to evaluate is by saying something which is a literal and palpable falsehood, as in (5) of my original list.² Evaluation via exaggeration is a subspecies of this way to evaluate. Consider these examples:

Someone says of a university, "It's a rest-home for old ladies."

Someone says of Mr. Smith, "He's not a man; he's a mouse."

What leads people to tell these palpable lies? For notice, they do not get away with it; no one believes them. But of course they were not trying to get away with it. In fact, were they to get away with it—that is, were their statements believed at the literal level—the point of these utterances would be lost. One only succeeds in evaluating a seat of higher learning by referring to it as a rest-home for old ladies when that is *not* what it is; we do not, as a rule, evaluate an actual rest-home for old ladies by saying that it is one.

Exaggerations work in the same way. To say of Bob that he is ten feet tall, or even more that he always seems to be about ten feet tall, is often to evaluate Bob or his physique. We cannot say quite how he, or his physique, is being evaluated until we know more about the situation: (the actual speech act).

(B) We can, and often do, evaluate by stating a fact, as is the case with (6) in the original list. Thus we can say of a man that he is honest when this is true, or that he is kind to animals when this is true; and often to do so is to evaluate him or his behavior. However, the point made above, that it is acts not idioms which evaluate, is important here. When we state a fact about a person we may well be evaluating him, but we need not be doing so. That is, whenever we state a fact about a person (thing, event), whatever the fact, it is a contingent matter whether or not we are also evaluating him in that utterance.³ We have many ways to indicate whether we are evaluating some person when we state a fact (or say anything) about him, not all of them verbal; often, when an actual speech act is being considered rather than just a sentence, it is not necessary to indicate at all, i.e., the context, in the largest sense, does it. But it is worthwhile exemplifying the sorts of techniques available to us for indicating that a piece of fact stating discourse is,

² On a good many occasions when we utter a literal and palpable falsehood, we are speaking metaphorically. However, I prefer to speak of the function of literal and palpable falsehoods in evaluation rather than the function of metaphor in evaluation because, in a precise sense, I am very unsure what metaphors are.

³ There are at least near-exceptions to this general statement. See the fourth way to evaluate below (D).

in addition, an evaluation. For positive evaluations, the principal techniques involve intonation. But with negative evaluations there is a rich verbal repertoire. We can, for instance, attribute certain completely current word usages to others in the phrase "is what people call" or "so-called" ("... is a so-called intellectual"); or we can use the phrase "what used to be called" ("... is what used to be called sin"). We can tack "all right" or "I suppose" on to the original sentence ("He's an honest man, I suppose"); we can add almost any *but*-clause. We can place an intensifier, negative or positive, before the operative noun ("It's a damned horse"), add "too" before the operative adjective, and so on. However, it must be emphasized that, most of the time, these verbal tricks are quite unnecessary. Items of context, in the largest sense—particularly who is speaking, to whom, and, in what circumstances—are usually sufficient to determine whether the utterance is an evaluation.

(C) We can, and frequently do, evaluate by making a simile-statement, as in (7) in my original list. Consider the utterances:

He has a mind like the crack of a whip.
She looks like a princess.

How do princesses look anyway? Of course, princesses in stories are all beautiful, unless expressly designated as not being so, just as all witches are old and ugly, but the real princesses we have about today all seem to be more like thoroughbreds than beauties. Why then does it evaluate a woman's physical appearance highly to say she looks like a princess? Something, doubtless, to do with conventional beliefs and conventional attitudes. This needs sorting out.

To evaluate using simile one says that what is to be evaluated *is like* something else; or perhaps that it *looks like*, *feels like*, *tastes like* something else. But there is more to it than that; not every statement of this type can be used evaluatively, anyway in our society or outside special contexts. To say of some dish that it tastes like potted donkey is, in our society, to evaluate it negatively. But this is not necessarily true in all societies (consider the French-English differences over snails and frogs as items of diet). Nor is it true for all similar simile statements. If I say of some otherwise unidentifiable dish placed

in front of me that it tastes like roast veal I shall not (without in addition making a face, or unless it is known that I hate veal, or similar) have succeeded in uttering evaluatively. To say of some dish that it tastes like roast veal invites the reply that it is, in fact, lamb: or the reply that I am right, it is roast veal; where to say it tastes like potted donkey invited no reply of this type. Such an utterance succeeds in evaluating (if it does) because of attitudes prevalent in our society over donkey as an item of diet.

To summarize, we perfectly well evaluate using a simile statement. However, such evaluation only comes off if the people hearing the utterance either have, or are aware that others have certain attitudes toward, or (conventional) beliefs about, what appears in the predicate of the simile statement. The matter is in fact more complex than that but the approximation will do for my present purposes.

(D) We can, and often do, evaluate by using a gerundive expression.⁴ Such an expression is *prima facie* evaluative, as with (8) in my original list. This is the most complicated way to evaluate so far considered and needs more sorting out than the others.

When we say someone is tall or a coward, we attribute to him a certain characteristic, though the characteristics which can be attributed in this way can be of a great many types. Most of the time when we do this, and when the utterance is evaluative, it is evaluative in one of the ways already noticed. However, some predicates, grammatically indistinguishable from the others, are different. They have what I call the Ziffian characteristic. This must now be explained.

In his book Paul Ziff⁵ was theoretically trying to find the meaning of the word "good."⁶ He reached the conclusion that what the word "good" means is: *answering to certain human interests*.⁶ This would appear to be obviously false. That is, when someone says "X is good," he is not *saying* that X answers to certain human interests, though it is doubtless true that, if it is correct or justifiable to say that X is good then X being the way it is answers to certain human interests.

I am not now interested in discussing "good" but the Ziffian characteristic is important in more areas than just for "good." With certain adjectives it is

⁴ The term comes from P. Nowell-Smith, *Ethics* (Oxford, 1957).

⁵ *Semantic Analysis* (New York, 1960).

⁶ I rely here on Ziff's explication of "answering to certain interests." To do so is not a serious omission for reasons given below.

the case that, when the word is correctly or justifiably applied, what it is applied to answers to certain human interests. Thus, if we symbolize such a word by *W*, to say correctly or justifiably "*X* is *W*" involves that *X* answers to certain human interests. So far, none of this is very startling. It is a general empirical truth that when *any* adjective can be correctly or justifiably used in this way, then *X* answers to certain human interests (or disinterests). That is, it is a general empirical truth that we only remark, and have words to remark, items and properties when they answer to certain human interests (or disinterests). However, with certain adjectives, it is not a contingent but a necessary matter that the correct use of that adjective answers to certain human interests or disinterests. Such adjectives have what I call the Ziffian characteristic.

Let us symbolize an adjective with the Ziffian characteristic by *Z*. Then, if someone says correctly or justifiably "*X* is *Z*," it is necessarily the case that *X* answers to certain human interests or disinterests. Examples of words with the Ziffian characteristic are "sublime," "admirable," "vile," "salutary," "genuine," "fine," "filthy" and so on.

This gives rise to another way to evaluate. When we predicate an adjective with the Ziffian characteristic we have used a predicate where there is a logical connection between the attitudes or interests of some one or some group and the fact that the predicate correctly or justifiably applies. Whenever such a predicate is used, there is a *prima facie* case that the speech act is an evaluation.⁷ But only, of course, a *prima facie* case. There are, as noted earlier, all sorts of ways we can apply an adjective with the Ziffian characteristic and yet take away its evaluative power. Thus,

The view was sublime, I suppose, but it bored me to tears.

He's just too damn admirable.

The problem with adjectives which have the Ziffian characteristic is that there is a complete spectrum effect and it is an area where idiosyncratic use is very common. It is important to notice it only

because there are certain words which exhibit it in a marked degree; not because it is a well defined separate way to evaluate.

This account of what might be called Ziffian evaluation is very far from complete. However, this is not important for my present purposes. Most discussions of evaluation in this century have been exclusively concerned with Ziffian evaluation. One of my aims in this paper is to put Ziffian evaluation where it belongs: as one way, and not a particularly important way, to evaluate; one which sits beside *many* other ways to evaluate. The idiosyncracies of Ziffian evaluation are not enormously important in the study of evaluation in general just because these idiosyncracies are what make it different from most evaluation. These points are discussed in more detail below.

These are by no means all the ways we evaluate⁸ but they comprise a fair enough cross-section to make at least some negative generalizations. However, before doing so, I want to return to the point made earlier that it is acts not idioms which evaluate. All the types of evaluation (as opposed to semi-evaluation) so far considered are accomplished by using sentences which look like descriptive sentences. Indeed, most of the sentences could be used for descriptive purposes as well as for evaluative purposes. There is nothing odd about this. The nature of the act we perform when we speak *of course* depends on the conditions under which it occurs. Thus, though there are verbal and vocal techniques to show that some utterance is an evaluation and how evaluative, most of the time this is quite unnecessary. To say *of an exhibit* in an art gallery "That's garbage" is pretty certainly to evaluate; while to say *exactly the same thing* in the same tone of voice of a bag full of scraps in the kitchen in answer to the question: "What's this?" is pretty certainly not to evaluate.

5. What has gone before is not just random talk. It is by means of such common sense inspection of actual evaluations that we can find out what is involved in evaluating. This is also the only way we can break down various inappropriate models by which, as philosophers at a point in time, we

⁷ One suspects that there is a logical connection here, but I can see no way to show that there is. I present it here as simply a fact about language.

⁸ Two notable omissions are evaluation using counter-factual conditionals and evaluation by giving (often impractical or unwanted) advice. For instance:

Of a child which is not one's own, one says, "If it were mine, I'd strangle it."

Of some painting one offers the advice, "Burn it."

But both these idioms have problems associated with them which are extraneous to evaluations as such and I have therefore not considered them.

find ourselves surrounded. To evaluate is not always or necessarily to place something on a scale or grade it, though we very well can evaluate things by placing them on a scale (when there is a scale available). It is not a matter of invoking or following rules, though, once some fundamental evaluating has been done, we perfectly well can evaluate in this way. Nor does evaluating something necessarily involve expressing emotions about or towards it, though using typically emotion expressing utterances is one way to evaluate. Nor does evaluating consist in stating our attitudes towards things, though we can evaluate by doing so; nor is it to state society's attitude towards things, though again we could evaluate in this way.

Every one of these ways to evaluate fails as a pattern for all evaluation. To be somewhat rough and ready: the rule following, grading, etc., explanations are no good as patterns because, as Urmson saw, the basis of such evaluation then remains totally obscure; clearly, in fact, such a basis (a set of standards) must be the result of evaluation. The expressive, emotive explanations of evaluation fail as patterns because it is quite certain that we can evaluate while experiencing no particular inner state suitable for expression. The attitude—or reaction—reporting or—expressing explanations fail as patterns because, quite simply, they eliminate evaluation. More importantly, all these expressions fail as models or patterns for all evaluation because, as I hope I have shown, they just do not fit—they do not even begin to fit—the sorts of ways we do evaluate every day: we do not, always, evaluate with reference to a set of standards; we do not, always, express anything when we evaluate; we do not, always, state a fact about anything, attitudes or anything else, when we evaluate.

I shall try now to give some sort of summary of the account so far given of evaluation.

I. *Semi-evaluatives*: We can evaluate something by making a statement about ourselves where the fact that it is offered as a fact about ourselves carries evaluative import. Thus, "It makes me sick," "I don't like it." Or we can say, "Ough!" or curl our lip.

And so on. Schematically:

(A) *Exclamations*
Ough!
Wonderful!

(B) *Statements of reaction*
It makes me feel sick
It makes me want to laugh
(C) *Statements of attitude*
I like it
It appeals to me

In these cases, evaluation is susceptible of some sort of emotive or attitudinal analysis in this sense: that the analysis depends on these sorts of sentences having a primary emotion-, attitude- or reaction-stating or expressing role.

II. *Evaluatives*⁹: We can evaluate something with respect to some property it has or with respect to something else. This can be done in a literal falsehood, an exaggeration, a statement of fact and so on. Schematically:

(A) *Literal falsity and exaggerations*
He's not a man; he's a mouse
Bob is ten feet tall
(B) *Fact stating*
He's a liar and a thief
Smith is very generous
(C) *Simile statement*
It tastes like potted donkey
It tastes like champagne
(D) *Ziffians*
It's sublime!
It's awful!

Let us suppose that *X* is evaluated with respect to some *Y* in such an utterance. *Y* may be important because it has certain properties in marked degree, like the crack of a whip or a grasshopper. In either case, whether what is important is things or properties of things, if the evaluation is to come off, then the members of the society in which the evaluation is used must have, or must know of, certain attitudes towards the things or properties in question; anyway suitable attitudes towards these things or properties under certain circumstances.

With all the items in the "evaluatives" list it must be emphasized that it is perfectly possible that any of the examples given could be so uttered as not to be evaluations or not fit into the categories they appear in above. For instance, "He's a fool" could be uttered under circumstances such as to fit under (A) or (B), and there is a case that it should go under (D). This difficulty with classification is not something which should surprise us. It is a general

⁹ The distinction between semi-evaluatives and evaluatives needs more justification than I have given it. However, as this is quite a complicated matter and I make no use of the distinction in my main argument, I do not pursue it here.

empirical fact about language that the majority of sentences can be used to perform many different, and different types of, speech act.

6. Clearly attitudes, in one way or another, play a large, though not a universal, part in the account of evaluations so far given. Other things, like reactions, or conventional beliefs, or supposed conventional beliefs (that all princesses are beautiful), also play a part. However, attitudes are both the most pervasive of these things which underlie evaluation and doubtless the most controversial. It is attitudes, therefore, that I shall discuss. However, it should be noticed immediately that though attitudes underlie much evaluation in my account they do not do so in any one way: there is no *one* relationship between evaluations which in some way involve attitudes and attitudes.

In giving an account of attitudes we must bear in mind that analysis must stop somewhere, preferably somewhere where the view is clear, and what is encompassed under the English word "attitude" seems to me to be such a point. That is, we have surely explained why a particular utterance of "He's a fool" was a negative evaluation (supposing it was) if we say (and it is true) that the audience to this utterance, and the speaker, have strong negative attitudes towards those of low intelligence; no further explanation of the attitudes involved is possible which does not become a lesson in English or an independent exposition in the philosophy of mind. My account of attitudes, therefore, will consist almost entirely of reminders of what we already know.

Attitudes do not have to do exclusively with physical action or stance (and never have: from the Latin *aptitudinem*) but with expressed states of mind. When it is not stance which is being referred to, attitudes are not momentary things, though they can change. They are prior to evaluation: animals and very young children have them. This is not strange. In virtue of being sentient animals, we all have attitudes towards pain and hunger and cold. Attitudes, at this basic level, are not things we have to learn about or learn how to express; they are things that we have automatically in virtue of the human condition. Highly sophisticated adults may have surprising attitudes—as, for instance, favorable attitudes towards pain—but these are in all cases odd; no baby is a masochist. We do not have to be trained or persuaded into our basic attitudes towards things and situations and conditions. But, of course, we can learn new attitudes over and

above the ones we have in virtue of our condition, both consciously and unconsciously. So there is, I would like to say, nothing surprising about attitudes; neither in that we have them nor in their diversity. It is, I suppose, a contingent fact that we do have attitudes at all—computers do not have them—but our having attitudes is fundamental to human life as we know it.

One further point; I have not considered above whether an evaluation can be true or false. I have not done so because that is a paper in itself, the main subject of the paper being the very diverse ways in which "is true" is used. Roughly; in so far as "is true" is, as P.F. Strawson¹⁰ once claimed an agreeing, approving predicate, evaluations will take the assessment "is true," along with many other assessment predicates (we can, of course, evaluate evaluations). But evaluations are not true or false the way simple, ordinary, paradigm factual statements are true or false; the Austinian analysis of truth (in "Truth") does not even plausibly cover "is true" applied to an evaluation. We should notice in this connection that the assessment "That is a true evaluation" is unnatural, though some use could doubtless be found for it; what we say more naturally is that an evaluation is fair or just or careful.

7. In my earlier talk of evaluation, I deliberately excluded the philosopher's paradigms of evaluative words for reasons given. Surely, however, it would be unreasonable to continue in this course indefinitely. My argument in the present section is to the effect that it is *not* unreasonable to continue in this course. I shall use "good" as my stalking horse.

The term "good" is complicated in its functions (in large part, at least, because it has the Ziffian characteristic) but it is in no sense essential for evaluation. This point can be seen easily enough. If some totalitarian power decreed that "good" should never again be used it would not hamper our ethical or practical talk seriously, and not because we have a synonym available or could invent one easily. The point is that there are a great many evaluative techniques which do not involve words having the Ziffian characteristic and all these would still be available to us. However, if a totalitarian power decreed that there should be no more evaluation, then the situation is all but unimaginable; not only would ethical talk cease but, if the decree could be rigorously enforced, the whole business of human life would come close to breaking down.

¹⁰ "Truth," *Analysis*, vol. 9 (1949), pp. 83-97.

If my case in the paragraph above is granted, we really need no further discussion of "good" in the present context. However, it is worth making two additional and interconnected points.

(A) To say of anything that it is a good *X* can sometimes be a simple factual statement and be used as such. That is, it is, on occasions, possible to claim that something is, say, a good Jersey bull or a good poodle and be, quite definitely and simple, right or wrong. Similarly, if someone says that Jones got a good degree when in fact he got a First, the statement is true. None of this is in conflict with the fact that "good" has the Ziffian characteristic; there is nothing about that characteristic which requires that words having it must be used in some sense non-factually.

(B) As a corollary to this point, the word "good," used quite normally, need not be used to evaluate or commend. That is, *of course* the speaker need not be commending Jones or his degree when he says, truly or falsely, that Jones got a good degree. Suppose you are asked to give a publisher information about Jones for the "About the Author" piece on a book jacket. Jones, you say, was born in 1918, attended Stowe School from 1931 until 1935, got a good degree at Oxford in 1938, fought in the

desert, married Jane Smith in 1946, and so on. These are just the relevant facts about Jones; there is not necessarily any evaluating going on.

Many philosophers have assumed in the past that if they really got straight about the use of "good" they would have made some serious advance in the study of evaluation and of ethics. I have tried to show that "good" is not essential for evaluation and is not necessarily evaluative. Furthermore, because it has the Ziffian characteristic, evaluations containing "good" are not of a type with all evaluation. Thus, though the investigation of the logical properties of "good," and the investigation in general of words with the Ziffian characteristic, may be a task of the greatest intrinsic interest, there is no reason to suppose it will get us to the kernel of evaluation or the kernel of ethics.¹¹

8. To summarize: I have here offered a partial phenomenology of evaluation. This is, without much doubt, worth doing in its own right. However, two important points came out immediately: (1) that it is acts (mostly, but not exclusively, speech acts) and not idioms which evaluate; (2) that the sorts of models we have been offered for all evaluating in this century do not fit, do not even begin to fit, the facts of our actual evaluations.

University of California, Santa Barbara

Received April 13, 1967

¹¹ There is a further defense which could be offered for the thesis that the study of "good" is what is important for ethics: this defense would consist in claiming that *only* using "good" can we evaluate *morally*. But as this claim appears to be obviously false, it need not delay us.

VIII. ON NAMES: SENSE AND REFERENCE

JOSEPH MARGOLIS

ONE of W. V. Quine's attempted economies in *Word and Object*¹ is that of the "reparsing of names as general terms" (p. 180). There is an important gap in the argument favoring the substitution, to which I wish to draw attention, that bears on the clarification of the function of proper names. Quine's proposal is given in summary in the following:

The equation " $x = a$ " is reparsed in effect as a predication " $x = a$ " where " $= a$ " is the verb, the ' F ' of " Fx ". . . "Socrates" becomes a general term that is true of just one object, but general in being treated henceforward as grammatically admissible in predicative position and not in positions suitable for variables. It comes to play the role of the ' F ' of " Fa " and ceases to play that of the ' a '. . . "Pegasus exists" becomes " $(\exists x) (x \text{ is Pegasus})$," and straightforwardly false; "Socrates exists" becomes " $(\exists x) (x \text{ is Socrates})$," with "Socrates" as general term, and probably true (with timeless "is," of course). "Socrates" is now a general term, though true of, as it happens, just one object; "Pegasus" is now a general term which, like "centaur," is true of no objects. The position of "Pegasus" and "Socrates" in " $(\exists x) (x \text{ is Pegasus})$ " and " $(\exists x) (x \text{ is Socrates})$ " is now certainly inaccessible to variables and certainly no purely referential position, but only because it is simply no position for a singular term; " $x \text{ is Pegasus}$ " and " $x \text{ is Socrates}$ " now have the form of " $x \text{ is round}$." (P. 179.)

The leading consideration is this. As Quine himself stresses, singular terms cannot occupy a predicative position (cf. also, pp. 95-100). But proper names are singular terms; even "Pegasus" counts as a singular term though true of nothing" (p. 95). It follows, therefore, that the "Socrates" of "is Socrates" (reparsed) and the "Pegasus" of "is Pegasus" (reparsed) are *not* proper names. What are they? It is difficult to say because of another consideration, namely, that proper names are said to have a reference but not a sense, *in the sense* that one cannot define a proper name, or state its meaning, or provide a synonym for it.²

Quine himself, on an earlier occasion, has said that talk about meanings essentially concerns "the *having* of meanings, which is significance, and *sameness* of meaning, or synonymy. What is called *giving* the meaning of an utterance is simply the uttering of a synonym, couched, ordinarily, in clearer language than the original."³ But if "Socrates" as a name has a reference but not a sense (in the sense specified) and if "Socrates" in the phrase "is Socrates" (reparsed) is a general term that has a sense, the connection between these seemingly identical, but in fact quite different, terms is not at all clear.

The difficulty is that "is Socrates" and "is Pegasus," construed as predicative expressions, require a sense but apparently have none as they stand. The reason is that "Socrates" and "Pegasus" (which are names), from which they are apparently formed, have no assignable sense that can be carried over to the allegedly general terms "Socrates" and "Pegasus." On this count, the predicative expressions are meaningless. I do not wish to deny, of course, that proper names may quite naturally devolve into general terms by way, for example, of connotations increasingly regularized in apparently predicative expressions like ". . . like Quisling," ". . . as Rembrandt might." But three things need to be remembered: (1) It is the regularized associations with the object and *not* the mere name that invests the derivative general term with a sense—a proper name cannot *qua* proper name be reparsed as a general term; (2) The derivative general term now enjoys what Quine calls "divided reference" (cf. pp. 90-95) and may be true, as he says "of each, severally, of any number of objects" (p. 90ff)—hence, must have a sense independent of the original name, in virtue of which it may be used predicatively (even of the original object named). (3) The properties associated with the derivative term are, characteristically, striking or prominent but not unique properties.

¹ Cambridge, Mass., 1960. References are all to the paperback edition published in 1964.

² Cf., for example, the view of John R. Searle, "Proper Names," *Mind*, vol. 67 (1958), pp. 166-173.

³ "On What There Is" in *From a Logical Point of View* (Cambridge, 1953), p. 11.

The upshot is that Quine must be mistaken in holding that, if "Pegasus" is construed merely as a general term " $(\exists x) (x \text{ is Pegasus})$ " is "straightforwardly false." For, if it is truly a general term, then if there is *anything* that is Pegasus (whatever sense be assigned to this term) the sentence is true. The only reason Quine could be so confident that it is false is because he claims that "Pegasus exists" is false. He *couldn't* know that " $(\exists x) (x \text{ is Pegasus})$ " is "straightforwardly false"—which shows at a stroke that the one sentence cannot be satisfactorily substituted for the other.

The "Pegasus" of "is Pegasus" (repared) cannot be a proper name. What is it, then? I suggest, solely for purposes of elucidation, that, as one possibility, it is "named 'Pegasus,'" or "christened 'Pegasus,'" or some such thing. " $(\exists x) (x \text{ is Pegasus})$," then, must be an ellipsis for " $(\exists x) (x \text{ is named 'Pegasus'})$ " or the like. But "named 'Pegasus'" will, on Quine's view, be one of a range of "indissoluble general terms" (p. 179), which makes it quite clear that it is not related to the singular term "Pegasus." For, the name "Pegasus" cannot be a part of an "indissoluble general term"; furthermore, if "Pegasus exists" is true then either " $(\exists x) (x \text{ is named 'Pegasus'})$ " is probably true or is true because it is already known that "Pegasus is named Pegasus" is true. So, there being something named Pegasus is not a sufficient condition for Pegasus' existing (what predicative condition could be?): there is no logical connection between the particle "—'Pegasus'" and the proper name "Pegasus," although there is a logical connection between the two sentences concerned. Consequently, *on the interpretation given*, far from " $(\exists x) (x \text{ is Pegasus})$ " being "straightforwardly false," it is very likely true and, more important, does not, if true, at all concern the truth of "Pegasus exists."

The reason for Quine's persistence here is his conviction that "nothing stands in the way of our making a clean sweep of singular terms altogether, with the sole exception of the variables themselves" (p. 185). But he is mistaken: he has not managed to eliminate proper names. He holds that "it can be shown that everything that used to be demonstrable or deducible from given premisses when 'Socrates' was manipulated unquestionably as a singular term is still demonstrable or deducible from those same premisses with the added help of the uniqueness premiss ' $(\exists y) (y \text{ is Socrates and Socrates only})$ '" (p. 187). However, construe "Socrates" here as "named 'Socrates'." The

uniqueness premiss thereby is rendered either false or probably false. Construe it as uniquely singling out the intended object, Socrates, when used in the context of a sentence. The phrase "is Socrates" cannot then be interpreted predicatively but only in terms of identity: the singular term remains. Fill out the predicate "is named 'Socrates'" more completely so that it may single out the intended object, Socrates. The paraphrase will depart radically from the sense of "Socrates exists." Furthermore, it will make no difference; it could not be reliably expected, because of the divided reference of general terms, to be true—the rest of the predicate expression "and *y* only" is then gratuitous and hearkens back to the undivided reference of singular terms. In effect, although "and *y* only" is the intent of the would-be paraphrase "is named 'Socrates,'" it is redundant; for, if Socrates *were* uniquely singled out by "is named 'Socrates'" (or by another predicate), the intent would be fulfilled; and if he were not, the paraphrase would be unsuccessful. *Nothing is added predicatively* by the phrase "and *y* only" here: it is the expression of a gamble, not the addition of a qualification. Consequently, contrary to what Quine says (on the interpretation given), it is not true that "everything that could be done with ' $(\exists x) (\dots x \dots)$ ' can still be done, in translation, on the premiss ' $(\exists y) (\dots y \dots \text{ and } y \text{ only})$ '" (p. 187). In fact, sentences of the first form may easily be true while the corresponding sentences of the second form are false. For, if it is true that Socrates exists, it may well be false that there is one and only one thing named "Socrates" (or one and only one thing satisfying any given condition). We must remember that the paraphrase *sans* the uniqueness condition provides neither a necessary nor a sufficient condition of the truth of "Socrates exists." But if this is so, then singular terms have by no means been eliminated. Also, Quine would not have demonstrated that "it would be pointless to defend either the singular term or the general term as the regular counterpart of the name "Socrates" of ordinary language" (p. 189). Everything points to the claim of the singular term (though not as Quine has thought to define it).

I draw attention to the fact that Quine is explicitly concerned to paraphrase "the name 'Socrates' of ordinary language." In so far as he is attempting to do this, the foregoing argument holds. Actually, there is another direction in which his attempted paraphrases move. For he says, meaning to resist the suggestion, that "there is a

feeling that in reparsing the names as general terms we forfeit part of their meaning, viz., the purport of uniqueness" (p. 182). Elsewhere, to capture this uniqueness (that he requires of his predicates, reparsed from names) he speaks of "the *ex hypothesi* unanalyzable, irreducible attribute of being Pegasus, adopting, for its expression, the verb 'is-Pegasus' or 'pegasizes'."⁴ From this point of view, it is true that the uniqueness condition of the paraphrases cannot be eliminated: $(\exists y)(\forall x)(x \text{ is Socrates if and only } x = y)$ is precisely what he requires (p. 186). It is also clear that, given this paraphrase, "Socrates exists" and the corresponding uniqueness sentences will be materially equivalent. But the difficulty now lies elsewhere. For, we have no idea *how* to specify *that* property that may be uniquely ascribed to Socrates and that property that may be uniquely ascribed to Quine, and so on for *everything* in the world. It does indeed look as if "the employment of terms like 'pegasizes' . . . does nothing more than introduce 'proper names' at the predicate level."⁵ If we construe the predicate "is-Socrates" otherwise (along the lines of "is named 'Socrates'"), it is, as has already been argued, quite likely that when "Socrates exists" is true, the corresponding uniqueness sentence is false. But we cannot, of course, argue that if the uniqueness sentence (on our construction) is false, then it is false that Socrates exists. So either the paraphrase must be rejected as inadequate or, if it is adopted as adequate, by definition, the sense of the predicate on which it turns remains an utter mystery.

The irony is that a proper name, that has (in the sense specified) no sense, yields a general term, that has a sense but one that is utterly unfathomable. Furthermore, although, by using proper names, particular things may be uniquely identified, we do not know, nor need to know, nor typically believe, nor need to believe, nor know straightforwardly how to determine, nor believe we know how, nor need to know how, nor need to believe we know how to determine, that *anything*

and *everything* has unique properties. Quine seeks here to replace the uniqueness of reference (what Strawson usefully amends in terms of "identifying" uniquely)⁶ with unanalyzable, unique properties. The truth is that it is not the name but the use of the name that provides for the uniqueness of reference. Consequently, under the circumstances, the attempted paraphrase is doomed. We should not even know, on inspection, whether "Socrates is-Socrates" is true. To single Socrates out, by name, is not to appeal to unique properties. On the other hand, if Quine's paraphrases were allowed, he would not have eliminated singular terms, since to apply his general terms predicatively to particular things, we should require singular terms to identify this or that particular to which the predicates would apply.⁷ Finally, if the general term "Socrates" *did* specify some unique property of Socrates, the phrase "and *y* only" of the proposed paraphrase would be redundant. If it is not redundant, the property specified is not the property of being uniquely Socrates (whatever that means); thereupon, the general term "Socrates" would behave entirely like the general term 'is named 'Socrates'.' What Quine has confused here is the notion of "the unique property of being Socrates" (which may be inescapably Platonistic) and some property (defined without reference to Socrates) that Socrates possesses. The first is dubious and, in any case, renders "and *y* only" superfluous; the second is normal and renders the paraphrase unacceptable (for reasons already given). There are no other alternatives.

Consider now sentences that avoid the use of "exists" (since for Quine, "the only way we can involve ourselves in ontological commitments [is] by our use of bound variables").⁸ "Pegasus flies" or "Cerberus barks," where "Pegasus" and "Cerberus" are singular terms, *is true*. Quine denies this, since the "singular term . . . fails to name."⁹ But the error is Quine's: "there really is a figure in Greek mythology whose name is 'Pegasus'."¹⁰ Consequently, speaking of fictional and mytholo-

⁴ *Ibid.*, p. 8. Compare here, Bertrand Russell, *Human Knowledge, Its Scope and Limits* (New York, 1948), Pt. II, ch. iii and Pt. IV, ch. viii, for some interesting parallels.

⁵ Herbert Hochberg, "On Pegasizing," *Philosophy and Phenomenological Research*, vol. 18 (1957), p. 553.

⁶ P. F. Strawson, "Singular Terms, Ontology and Identity," *Mind*, vol. 65 (1956), pp. 433-454.

⁷ *Ibid.*

⁸ "On What There Is," p. 12.

⁹ *Methods of Logic* (New York, 1959, revised ed.), p. 221. Cf. *Word and Object*, p. 182.

¹⁰ Cf. Leonard Linsky, "Reference and Referents" in *Philosophy and Ordinary Language*, ed. by Charles E. Caton (Urbana, 1963). Cf. also P. F. Strawson, "On Referring," particularly as amended in *Essays in Conceptual Analysis*, ed. by A. Flew (London, 1956), where he speaks of secondary referring uses.

gical entities does not commit us to the awkwardness of "truth-value gaps" for which Quine has constructed his paraphrase (pp. 177-179). Quine has hoped to "rephrase" or "translate" or "convert" "Pegasus" into a singular description and, *à la* Russell, to eliminate such descriptions.¹¹ But, since proper names have a reference but not a sense (in the sense already given), to understand a sentence employing a proper name one must know what is referred to by the use of that name; otherwise, clearly, questions of truth and falsity cannot arise at all. And if " $(\exists x) (x \text{ is Pegasus and } x \text{ flies})$ " is, on Quine's analysis, false, it must surely be rejected as a paraphrase of "Pegasus flies." Also, to know that "Pegasus" fails to name is, trivially, to know that no unique predicate can be formulated that could single out what "Pegasus" may have been supposed to name. But there is no relevant unique predicate then that could even be formulated. For names that fail to name (for which Quine wishes to provide), there simply is no predicative paraphrase possible. Divided reference and questions of relevance disqualify the maneuver.

I have emphasized that, in a sense specified, proper names have a reference but not a sense. One of the benefits of this insistence is the ease with which it eliminates the problem of truth-value gaps. For, a sentence purporting to predicate something of an object named cannot be judged in terms of truth unless we understand what the referent is of which something is being said.¹² Because proper names have no sense (in the sense specified), the question of truth-value cannot always arise for the relevant sort of sentences and arises only on the condition of grasping the name's referent. But *if* names had a sense (in the sense specified), the problem of truth-value gaps would be generated for sentences using names that did not name anything. Here lies the temptation to "translate" or "rephrase" a proper name as a description and then to eliminate the gaps in the manner of Russell's theory of descriptions.¹³ The difference between proper names and demonstratives like "this" strengthens the gain, since demon-

stratives *do* have a sense (however reduced) that may be given independent of their reference. Demonstratives, therefore, and definite descriptions do raise the specter of truth-value gaps, though the resolution of the problem, depending as it does again on reference, must follow the solution for proper names.¹⁴ There cannot possibly be any tactic for assigning truth-values to sentences used referentially that restricts its attention to either considerations syntactical or semantical (in the sense of "having a sense" specified).

A further advantage of our account is that proper names cannot be construed as "incomplete symbols," which can be paraphrased only (and satisfactorily) in the context of the statement in which they are used referentially. This goes against Quine again: names cannot be paraphrased at all.¹⁵ Quine mistakenly insists that "there is a gulf between *meaning* and *naming* even in the case of a singular term which is genuinely a name of an object"; and he criticizes his fictional opponent for having "confused the alleged *named object* Pegasus with the meaning of the word 'Pegasus' . . ."¹⁶ Ironically, Quine adopts the error of his opponent, the distinction between the sense and reference of proper names, but resolves the difficulty about truth-value gaps differently: the one invents a spurious ontology and the other assimilates names to singular descriptions.

Quine's program for an extensional language, then, requires the elimination of singular referring expressions. But proper names present a double difficulty: for one thing, not having descriptive sense, proper names cannot be assimilated to definite descriptions or even to demonstratives; for another, any extensional paraphrase presupposes, what is not obviously true and what constitutes a logical (not an epistemic) claim quite independent of that regarding the purely logical role of any and all predicates (divided reference), namely, *that every thing that there is must have unique properties*. In a word, not having descriptive sense, names cannot denote anything though they may name things and though they may be used, in using sentences

¹¹ "On What There Is," pp. 7, 12.

¹² Cf. Gottlob Frege, "On Sense and Reference" in *Translations from the Philosophical Writings of Gottlob Frege*, ed. by Peter Geach and Max Black (Oxford, 1952). The original paper, "Über Sinn und Bedeutung," appeared in *Zeitschrift für Philosophie und Philosophische Kritik*, vol. 100 (1892), pp. 25-50. Page references to the German original are given with the English translation cited, following a slash.

¹³ "On What There Is," p. 7.

¹⁴ Cf. P. F. Strawson, "On Referring," *Mind*, vol. 59 (1950), pp. 320-344.

¹⁵ "On What There Is," pp. 6-8.

¹⁶ *Ibid.*, p. 9.

to assert and the like, to refer to things. Furthermore, though terms may denote the members of a class of things, doing so does not presuppose that each such member possesses unique properties.

There is one final advantage that I wish to claim for the view that emphasizes that proper names lack a sense (in the sense specified) but have a reference. Frege posed the question of how to account for the difference in cognitive value between two true identities, " $a = a$ " and " $a = b$," that is, where " a " and " b " have the same referent.¹⁷ Unless, therefore, " a " and " b " differed in sense, so he argued, the difference could not be accounted for. Also, it is important to notice, in asking his question, Frege clearly took the difference in cognitive value to be connected with the difference between analytic and synthetic statements. The difficulty is that, in clarifying the difference between sense and reference, Frege draws his illustrations from descriptive phrases, that have both sense (descriptive sense) and reference.¹⁸ Furthermore, in explicating the sense of a proper name, Frege holds that it will be "grasped by everybody who is sufficiently familiar with the language or totality of designations to which it belongs"—and, "in the case of an actual proper name such 'Aristotle' opinions as to the sense may differ." "It might, for instance," he says, "be taken to be the following: the pupil of Plato and teacher of Alexander the Great."¹⁹ But, of course, on *this* basis, Frege *could* not distinguish between " $a = a$ " " $a = b$ " as between analytic and synthetic statements. "Aristotle is Aristotle" and "Aristotle is the teacher of Alexander the Great who was born in Stagira" might then both be analytic. The change from analytic to synthetic status would depend, in all cases, on the "opinions" of particular speakers. And, in fact, the sense of proper names would be bound to vary from time to time, from occasion to occasion, from speaker to speaker.

Nevertheless, *independent* of whether or not " $a = a$ " and " $a = b$ " are analytic, we should wish to say that, if true, they have differing cognitive values. But since their reference is the same, the difference cannot be explained in terms of this but must, in some sense, have to do with the fact that " a " and " b " are different signs (to use Frege's

term). What is proposed is that the difference in their cognitive value be assigned to the fact that in the first identity the same sign is employed twice and in the second, two different signs are employed. Consequently, *if true*, the two identities differ in cognitive value, whether analytic or not and independent of "opinions" about the "designations" that belong to the sense of the name. Here, the type-token distinction will suffice.²⁰ The information we have in identities that employ *only* proper names *must* be information about the names only, that (in " $a = b$ " for instance) two different names have the same reference. The information we have in identities that employ only descriptive phrases cannot be said to be about names or words only. Consequently, in the second case, *but not in the first*, the question of whether the statements concerned are or are not analytic affects the question of a difference in cognitive value. But here, once again, we see a strong reason for resisting the assimilation of proper names to descriptions.

Also, although "Aristotle was the teacher of Alexander the Great" and "The author of the *Poetics* was the teacher of Alexander the Great" are materially equivalent, they are not synonymous or translations one of the other. The reason again is that "Aristotle" has a reference but not a sense. And this distinction can be preserved whether or not we propose, along Fregean lines, to judge the pair as analytic or synthetic. The equivalence of the two depends once again on their having the same reference; but the analyticity of "The author of the *Poetics* was the teacher of Alexander the Great" depends on "opinions" of the sense of the two singular descriptions involved. Since two descriptions are not involved (or two expressions having a sense) in "Aristotle was the teacher of Alexander the Great," the question of analyticity *does not arise*; but, having decided it for the former statement, analyticity may be *assigned* to the latter to preserve a certain symmetry. The question in the present context is idle, however, since both the name and the descriptions are taken to refer uniquely to the same object. The identities will be equivalent, if true, *whether or not* "the author of the *Poetics* was the teacher of Alexander the Great" is analytic.

¹⁷ *Loc. cit.*

¹⁸ "On Sense and Reference," p. 57/26-27.

¹⁹ *Ibid.*, pp. 57-58/27-28.

²⁰ Cf. Searle, *loc. cit.*

Also, in explaining his view of sense, Frege says: "A proper name (word, sign, sign combination, expression) *expresses* its sense, *stands for* or *designates* its reference. By means of a sign we express its sense and designate its reference."²¹ Further, to clarify sense, Frege says that "the truth value of ' $a = b$ ' is the same as that of ' $a = a$ '" because "the reference of ' b ' is the same as that of ' a '" (with which the present account agrees); but to explain why "the two sentences do not have the same cognitive value," he holds that "the thought expressed in ' $a = b$ ' differs from that of ' $a = a$ '."²² And, of course the *thought* expressed in a sentence like " $a = b$ " will depend on the *sense* of " a " and the sense of " b "; and this, we have seen, will vary

depending on the "opinions" of people.²³ But since, on Frege's very view, the cognitive value of the two identities *may* be the same, since the "opinions" of people *may* assign the same sense; for example, to "Cicero" and "Tully," the thesis that proper names have a sense will not actually explain—what calls for explanation—the apparent difference in cognitive value of " $a = a$ " and " $a = b$." I conclude, therefore, that, only if the significant use of a proper name (call it "sense," if we must) is freed altogether from the notion of descriptive sense, will it be possible to account for differences in cognitive value between identities employing only names.

University of Toronto

Received May 22, 1967

²¹ *Op. cit.*, p. 61/63.

²² *Ibid.*, p. 78/50.

²³ *Ibid.*, 58n/28.

IX. THE INDISPENSABILITY OF TRUTH

HERBERT HEIDELBERGER

EVER since Ramsey published his paper "Facts and Propositions" it has been widely held that "there is no separate problem of truth"; by which it is usually meant that the word "true" is redundant and in consequence is eliminable in trivial fashion from any sentence in which it appears. Perhaps the clearest cases of such redundancy occur when truth is attributed to a sentence or proposition which is explicitly named. Thus, a man who says of the sentence "Snow is white" (or of the proposition which it expresses), that it is true does no more than to assert that snow is white; and in saying that it is false, he asserts its negation. More difficult for Ramsey's view are statements such as "The first sentence written by Plato is true" and "All consequences of true propositions are true," where the attribution of truth is to sentences or propositions that are described or otherwise indicated, but not named. Yet even here, Ramsey argues, "true" can be removed with no significant loss, although it is allowed that in these cases the undertaking requires a little ingenuity. This view of Ramsey's—defended in recent years, most notably by Strawson—has come to be known as the redundancy theory of truth and that is how I shall refer to it.¹

To appreciate the significance of the redundancy theory, it is important to see how radically it departs from some traditional theories of truth. From the perspective of the redundancy theory, such older theories as the correspondence and coherence theories are chasing after a will o' the wisp. However different they may be in other respects, these theories share the assumption that the concept of truth can be analyzed, that the predicate "true" can be defined. But if, as the redundancy theory has it, there is no genuine concept of truth, if the predicate "true" is indeed redundant, there is, of course, nothing that requires

analysis and nothing that needs to be defined. To be sure, on both traditional and redundancy theories, "true" can be eliminated. But according to these traditional theories this is because "true" can be replaced, replaced by something like "corresponds" or "coheres"; whereas on the redundancy theory "true" can be eliminated without being replaced.

In this paper I wish to show (1) that the redundancy theory cannot plausibly be formulated as some would have it in terms of sentences or statements but should be formulated, rather, as Ramsey had it in terms of propositions; (2) that there are sentences from which "true" cannot be eliminated without being replaced; and (3) there is good reason for thinking that if "true" can be replaced at all, then it can be replaced only by an expression which makes reference to what I shall call a Platonic relation.

By a Platonic relation I mean simply a relation between an abstract (or "platonic") entity—a proposition, quality, or relation—and something else which may be another abstract entity, but might as easily be a concrete thing or event. The terms "correspondence," "coherence," "exemplification," "instantiation," "belonging to," and "having" are among those commonly employed by philosophers to designate Platonic relations and many of the traditional theories of truth were formulated using these terms. Thus it has been claimed that a statement or assertion is true provided that it expresses, or is identical with, a proposition which corresponds to the facts; or with a proposition that is exemplified in some state of affairs; or with one that coheres with another body of propositions. And some have claimed that subject-predicate statements of the form "*a* is *F*" are true when *a* instantiates or has the quality *F*(ness), or conversely when *F*(ness)

¹ The theory is defended in many places. Among these: P. F. Strawson, "A Problem About Truth—A Reply to Mr. Warnock" in G. Pitcher (ed.), *Truth* (Englewood Cliffs, 1964); A. J. Ayer, *Language, Truth and Logic* (New York, 1936), pp. 87–90; A. J. Ayer, "The Criterion of Truth" in M. Macdonald (ed.), *Philosophy and Analysis* (Blackwell, 1954) pp. 237, 238; C. L. Stevenson, *Facts and Values* (Yale, 1963) pp. 214–220.

belongs to *a*. A similar account can be given of the truth of other kinds of statements. It can be said that a relational statement of the form "*aRb*" is true whenever *R*(hood) is exemplified by the ordered pair *a*, *b*; and a universal statement "All *F*'s are *G*" is true when whatever exemplifies *F*(ness) also exemplifies *G*(ness); and so on.

Strawson thinks that relational accounts of truth such as those given above are mistaken. "It is evident," he writes in his contribution to the symposium on "Truth" with John Austin, "that the demand that there should be such a relation is logically absurd; a logically fundamental type mistake."² The demand that Strawson thinks absurd is one that I think reasonable and one that I shall try to make out in the sequel.

I

The redundancy theory seems to admit of a straightforward formulation: A sentence '*S*' ("Snow is white") is true if and only if *S* (Snow is white); and a sentence '*S*' ("Snow is not white") is false if and only if $\sim S$ (Snow is white). (If we think we know what is meant by the expression "the use of a sentence" and conceive of statements as such "uses" we might wish to replace "sentence" in the formulation by "statement.") Now it happens that a good many philosophers who subscribe to the redundancy theory of truth also subscribe to another very familiar theory. I shall not attempt to describe this second theory except to say that it is a theory of referring according to which sentences (or statements) like "The present king of France is bald" and "All of John's children are asleep" are, in certain circumstances, neither true nor false. What is of interest to us about this familiar theory of referring is that although this theory and the redundancy theory of truth (as formulated above) are often held jointly, the two theories are, in fact, inconsistent. Proving this, I believe, is a fairly simple matter.

Let us choose some sentence (or statement) '*S*' which, according to this familiar theory of referring, is neither true nor false. Assuming the theory, we have

(1) '*S*' is neither true nor false,
from which assumption we can derive the conjunction

(2) $\sim ('S' \text{ is true}) \ \& \ \sim ('S' \text{ is false})$

which, when we make the inference sanctioned by the redundancy theory of truth,³ reduces to

(3) $\sim S \ \& \ S$

and this, in its turn, yields the explicit contradiction

(4) $S \ \& \ \sim S$.

As the conjunction of the two theories implies a contradiction, the theories themselves must be inconsistent.

It may be well to consider very briefly a pair of natural replies to this argument. It may be said in the first place, that the conclusion is not a contradiction as it consists of two sentences (or statements) neither of which possesses a truth value. If, however, we apply the redundancy theory to the conclusion we get

(5) $('S' \text{ is true}) \ \& \ ('S' \text{ is false})$

which—on any reasonable theory, including the aforementioned one—does have a truth value and is inconsistent with (1). Moreover, even if we stop at (4), and concede that it formulates no contradiction, it is surely a bad theory that leads us from statements like (1) and (2) which, we are assuming, are true to (3) and (4) which are neither true nor false. On the second reply, one would refuse to countenance the inference from (2) to (3) on the ground that '*S*' lacks a truth value. But to reply in this way is to give up the theory; for it is now admitted that we have a meaningful use of "true"—in $\sim ('S' \text{ is true})$ —to which the redundancy theory does not apply, i.e., we have a counterexample.

It is of some interest, I think, that the redundancy theory as stated is inconsistent with a certain theory of referring, if only because there are philosophers who embrace both. At the same time it should be clear that the above, conceived as an argument against the redundancy theory, does not need to rely on anything as controversial as a theory of referring. We can take any sentence that is without truth value, for example, "The class of all classes which are not members of themselves is a member of itself," or "Knock before entering," suppose it to be '*S*,' and the argument would go through. It is noteworthy that the redundancy theory conflicts with another theory, but it is still more important to note that the redundancy theory conflicts with something that is undeniably true: that there exist sentences that are neither true nor false. Indeed this is not the only truth with which the redundancy theory conflicts; I

² In Pitcher, *op. cit.*, pp. 36-37.

could cite others, but as the argument that I have given, if sound, is conclusive, there is no need to do so.³ Obviously, some modification is in order.

It may be thought that the conflict can readily be resolved by introducing a somewhat different conception of a "statement." If we think of a statement, as does Quine, as a sentence with a truth value we might attempt to define "'S' is true" as the conjunction of "'S' is a statement and S." This can be shown to be unsatisfactory. If our sentence is "Knock before entering" then the definiendum will read "'Knock before entering' is true," which is false; whereas the definiens will read "'Knock before entering' is a statement and Knock before entering," which is without truth value, containing as it does a conjunct that is without truth value. So that when we choose a sentence that is without truth value our definiendum will have a truth value, although our definiens will not. The trouble with the definition of "true" in terms of statements lies in its requirement that we not only mention, but also use the sentence to which we attribute truth. The requirement can be removed by the use of a well placed quantifier.

"'S' is true," let us now say, provided that "there is a proposition p which is designated by 'S' and p ."⁴ If "Knock before entering" is again our sentence, the definiendum will turn out false and so will the definiens, since (presumably) it is false that there is a proposition designated by "Knock before entering." Further, the inference from step (3) to step (4) in the previous argument can now be blocked. For that inference is sanctioned by the assumption that if a sentence is not true its negation is true, and that if a sentence is not false its negation is false, but on the amended definition of "true" it is possible for both a sentence and its negation to be neither true nor false. The

definition (which is one of Carnap's) solves our immediate problem.⁵ And although we have a definition, and in this respect something similar to the older accounts of truth, it might not unreasonably be said that once we have specified our proposition there remains no further—or "separate"—problem of truth. On the other hand, it is not hard to understand how a militant anti-Platonist would feel that the virtues of the redundancy theory have been compromised with the invocation of both propositions and the Platonic relation of designation. If, however, we wish, as I do not, to preserve the redundancy theory there seems to be no getting away from them. I shall return to this definition in Sect. III.

II

So far, I have tried to show only that in order to formulate a plausible version of the redundancy theory of truth, we have to have propositions. I want now to show that if we are to eliminate "true," we need to replace it by an expression referring to a Platonic relation. I shall consider occurrences of "true" where it is predicated of something other than the name of a proposition. It seems to be agreed by all parties that these are the crucial cases.

In "Facts and Propositions" Ramsey considered the statement

He is always right

for which he suggested as a paraphrase:

(p) If he asserts p then p .⁶

It is not clear whether Ramsey intended the last occurrence of ' p ' to fall within or outside the scope of the universal quantifier; either way, however, the paraphrase is unsuccessful. If ' p '

³ I shall cite one. The conflict is with the true statement, "There exists a meaningful sentence" is a synthetic statement." This is shown as follows. From "Snow is white" we deduce "The sentence 'Snow is white' is true," and then "The sentence 'Snow is white' is meaningful," which yields "There exists a meaningful sentence." By parallel reasoning the same conclusion can be drawn from "Snow is not white." Now any statement which is implied by each of two contradictory statements is itself an analytic statement; therefore given these arguments "There exists a meaningful sentence" is an analytic statement. But as it is evident that "There exists a meaningful sentence" is a synthetic statement and since the inferences other than from "Snow is white" to "The sentence 'Snow is white' is true" and from "Snow is not white" to "The sentence 'Snow is not white' is true" are unexceptionable, it must be that at least one of these inferences is mistaken.

⁴ Why not "There is a statement (in Strawson's sense) p which is made by 'S' and p ?" But it is not clear that statements as conceived in this definition are any less mysterious or problematic than a straightforward Platonic entity like a proposition. For so conceived they are not simply kinds of sentences, but unspecified entities of their own raising problems analogous to those raised by propositions. As we might ask of propositions "Can they exist without sentences to designate them?" ("or without persons to believe them?"), just so we could ask of statements "Can they exist without being made?" "Can they be made other than by sentences?" There seems to be no obvious way of answering such questions.

⁵ Cf. *Introduction to Semantics* (Cambridge, 1948), p. 91. A more recent formulation is in *The Philosophy of Rudolf Carnap* (La Salle, 1963), p. 901. What I shall say about the earlier definition will pertain, in an obvious way, to the later definition.

⁶ Pitcher, *op. cit.*, p. 17.

falls within the scope of the quantifier, then it is simply an isolated variable to which no predicate is adjoined. (I discuss this point at length in Sect. III.) If ' p ' falls outside the scope of the quantifier, it becomes an unbounded variable functioning much the way that the final ' x ' does in the formula: " $(\forall x) (Fx) \supset x$." The result is (at best) a sentence schema which cannot be a proper paraphrase of the original, as that is a sentence.

Strawson also attempts to rewrite sentences in accord with the redundancy theory, although he casts his paraphrases in more conventional English. Among the sentences that he considers is this one:

What he said about the house is true

for which he offers as a paraphrase,

There is something which he said the house was and which it is.⁷

The paraphrase seems accurate enough, but conceals, I believe, a Platonic relation. To appreciate this we have merely to ask ourselves what that something is which he said the house was and which (indeed) it is. What in general could that something be if not a property; a property which belongs to the house or which the house may be said to have, exemplify, or instantiate? Were we to write out Strawson's paraphrase in the logician's idiom it would, I suggest, be most naturally rendered as:

There is a property F (ness), such that F (ness) is what he said the house had (exemplified, instantiated) and the house has (exemplifies, instantiates) F (ness).

So the paraphrase eliminates "true," but only by replacing it with a reference to a property and an unwanted Platonic relation.

It is of course possible—albeit with some concessions to style and to grammar—to formulate the paraphrase without using "has," "exemplifies," or "instantiates." Suppose that he had said (truly) that the house was expensive. Then we might say that there is a property—being expensive—which he said the house was and which (indeed) it is. It matters little how we say it. What is significant is that we are assuming the existence of a house, that our quantifier ranges over properties and in this instance picks out the property of being expensive, and that we are asserting some relation to obtain between the house and this property, that relation

being of a kind which we have called a Platonic relation.

Worthy of consideration is Strawson's general formula for dispensing with "true" when that word is predicated of something other than an explicitly mentioned statement. Paraphrases that are faithful to the formula will make reference to the unmentioned statement by means of "a clause introduced by 'as' or by some other conjunction."⁸ Thus, for

(1) What the Pope says is always true
he offers

(2) Things are as the Pope says they are
and for

(3) What you're thinking isn't true

(4) Things are not as you think they are.

Again the paraphrases seem to be accurate; although "as" looks as though it might conceal a Platonic relation. We can determine whether it does by casting (1)–(2) and (3)–(4) in the more exacting mode of quantificational logic. It is easy to formulate them in that mode if we allow ourselves a Platonic relation and variables ranging over both propositions and that which, on traditional theories, makes them true. The latter we shall call "states of affairs," although "facts" or "situations" might do as well. Using (that p) (\exists that p) as universal and existential quantifiers ranging over propositions and ($\exists e$) as an existential quantifier ranging over states of affairs, for (1) we can write:

(that p) ($\exists e$) (The Pope says that $p \supset$ that p corresponds to (is exemplified by) e ;
and for (3):

(\exists that p) [You think that p and \sim ($\exists e$) (that p corresponds to [is exemplified by] e)].

Can these sentences be reworded so that no reference is made to the relation of correspondence or exemplification or any other Platonic relation? I do not see any way of doing so. Translations that might come to mind are

(p) (If the Pope says that p then p)
and

($\exists p$) (You think that p and $\sim p$)
but these, like Ramsey's paraphrase, leave us with variables unaccompanied by a predicate.

In light of the above it seems safe to say that Ramsey and Strawson have failed to make out

⁷ "A Problem About Truth—A Reply to Mr. Warnock" in Pitcher, *op. cit.*, p. 79.

⁸ *Ibid.*

their case; and indeed that the reasons for their failure lend some weight to the view (which I hold to be true) that if "true" is to be eliminated it can be eliminated only if it is replaced by an expression referring to a Platonic relation. But many philosophers would resist the view nevertheless because they believe that the kind of definitions of "true" that Carnap proposes in his various writings, especially in his *Introduction to Semantics* (Cambridge, 1948) are adequate definitions which do not invoke Platonic relations. I believe, on the contrary, that Carnap's definitions either invoke Platonic relations or are devoid of sense, and that is what I shall try to show now.

III

In *Introduction to Semantics* Carnap proposed three interestingly different definitions of "true." In the first of these he defined what he called an "absolute conception of truth"—truth as attributable to propositions that are explicitly mentioned.⁹

(That) p is true if and only if p .

To this somewhat limited definition, I can offer no objections.¹⁰ Carnap's second definition is for sentences of the subject predicate form:

A sentence a is F is true if and only if the designatum of a has the designatum of F .¹¹

We should note that here the definiens make reference to two kinds of Platonic relations: the semantic relation of designation holding between words and their designata (between ' a ' and the object a and ' F ' and the property F (ness) and) having which is a relation between the two designata (between a and F (ness)).

His third definition of "true" is applicable to any sentence:

A sentence ' S ' is true (in L) if and only if there is a proposition p such that ' S ' designates p and p .

The third definition—to which we have alluded earlier—has the advantage of simplicity over the traditional definitions of "true," as well as over Carnap's second definition. The present definition

makes use only of the semantic relation of designation, whereas on the previous definition we needed both having and designation. And were the traditional theories to give an account of the truth of sentences, presumably they too would have to make reference to two Platonic relations. One, a semantic relation such as that of designation or of meaning holding between a sentence and a proposition; and two, depending upon the theory being advanced, a relation of correspondence, exemplification, or coherence between the proposition and the state of affairs or set of propositions which would on these theories render the sentence true. So Carnap's third definition has the virtue of requiring one less Platonic relation than that of traditional theories; but the definition, I believe, is inadequate.

Carnap tells us that ' p ' in his formula is a propositional variable; therefore any constant with which we might replace this variable would be the name of a proposition. (Just as in " x is mortal," ' x ' can be replaced by the name of some particular thing, e.g., "Socrates.") But if ' p ' is replaced by a name in Carnap's definition of "true," the locution that results will be unmistakably ill-formed. For suppose that our proposition is that snow is white; then substituting in Carnap's formula we arrive at:

"Snow is white" designates that snow is white and that snow is white.

This locution is not, as might be expected, a conjunction of two sentences, but an incoherent joining of a sentence to a name: the sentence "Snow is white" designates that snow is white" and the propositional name, "That snow is white." What we have is comparable to the joining of "William Shakespeare is a poet" to "William Shakespeare," or the joining of "William Shakespeare is a poet" to the sentential name "Milton is a poet"; neither joining or forming a coherent English sentence. It is true that Carnap's definition looks a lot better than these examples would suggest, in fact the definition looks coherent. And the reason why we may be deceived into thinking that it is coherent, is because we tend to suppose that the last occurrence of ' p ' is functioning as a schematic letter, as it does in the formula

⁹ Carnap, *Introduction to Semantics* (op. cit.), p. 90.

¹⁰ Although it is doubtful that it is a definition. Cf. J. F. Thomson, "A Note on Truth," *Analysis*, vol. 9 (1949), pp. 67-72.

¹¹ Carnap, *Introduction to Semantics* (op. cit.), p. 24. Compare Sextus Empiricus: Now as to this definite proposition, "This man is sitting," or "This man is walking," they declare that it is true when the things predicted, such as "sitting" or "walking," belong to the object indicated. *Against the Logicians* (Cambridge, 1933), vol. II, p. 100. The definition is similar (as I understand it) to Austin's definition of "true" (cf. Pitcher, op. cit., p. 22).

" p or q ." We must remember, however, that it cannot possibly be functioning in that way, otherwise Carnap would be saying that the *bona fide* sentence "'Snow is white' is true" is logically equivalent to the sentence schema " $(\exists p)$ ('Snow is white' designates p) and p " which quite clearly he does not and should not want to say.¹²

We have then the same difficulty here as with Ramsey. If the last ' p ' functions as a schematic letter, that is if it can be replaced by a sentence, we get something that is well-formed but incomplete. On the other hand, if we take ' p ' to be a genuine variable bound by the existential quantifier then what we get is a stray variable, or a stray name if we make a substitution, with no predicate to attach to it.

It must be acknowledged that a number of quite respectable logical systems contain propositional variables without there being any predicates ostensibly attached to them. Thus, for example, on p. 184 of Lewis and Langford's *Symbolic Logic* (New York, 1959) there appears the formula: $(\exists p)(p)$. But the existence of such formulae by themselves has no bearing upon the issue under discussion. For it may be that when such formulae are interpreted a predicate will be supplied which is not written into the formula itself. Such, in fact, is the case with the formula mentioned above. In

interpreting his formula Lewis adds the predicate "true," reading it: "There exists at least one proposition p which is true" (p. 184). There is of course nothing that requires the addition of "true" rather than another predicate. But the point is that if we are to get an English sentence out of the formula we have to supply some predicate and that is what Ramsey and Carnap fail to do. Because their theory of truth precludes their adding any predicate whatever, they must leave us with formulae for which they have provided no sense.

There remains one final matter. Many philosophers in recent years have come to underestimate the problem of truth because, as I believe, they tend to overestimate the significance of Carnap's "absolute conception of truth." There is no denying, or at least I would not wish to deny, the following: If x is a sentence which is the result of filling the blank in "It is true that—" with another sentence y , then x is logically equivalent to y . An appreciation of this truth has led philosophers to say that "true" is redundant, or that "true" does not denote a property, or that it has a purely "performatory function." And for all I know one or more of these views does in fact hold of "it is true that." But from this, as I have tried to show, nothing of significance follows concerning "true" or "is true."¹³

University of Massachusetts

Received July 10, 1967

¹² The point made in this paragraph was anticipated by Wilfrid Sellars, although the conclusions he draws differ considerably from my own. See his important paper, "Truth and Correspondence" in his *Science, Perception, and Reality* (London, 1963).

¹³ I wish to thank Murray Kiteley, Roderick Chisholm, Michael Goldman, Jaegwon Kim, and Leonard Krimerman for the considerable help they have given me with this paper.

PRINTED IN ENGLAND
by C. Tinling & Co. Ltd., Liverpool, London and Prescott

AMERICAN PHILOSOPHICAL QUARTERLY

MONOGRAPH SERIES

Edited by NICHOLAS RESCHER

The *American Philosophical Quarterly* also publishes a supplementary Monograph Series. Apart from the publication of original scholarly monographs, this is to include occasional collections of original articles on a common theme. These monographs can be purchased separately but are available to individual subscribers at *half price* (though not to institutional subscribers).

No. 1. STUDIES IN MORAL PHILOSOPHY. Essays by Jerome B. Schneewind, Lawrence L. Haworth, David Braybrooke, Kai Nielsen, Jesse Kalin, Michael Stocker, and G. P. Henderson. May 1968 publication. \$6.00

No. 2. STUDIES IN LOGICAL THEORY. Essays by Montgomery Furth, Jaakko Hintikka, James W. Cornman, Robert C. Stalnaker, Alan Hausman and Charles Echelbarger, Ted Honderich, and Colwyn Williamson. July 1968 publication. \$6.00.